NPS-AM-12-C9P07R02-057

EXCERPT FROM THE

PROCEEDINGS

OF THE

NINTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM WEDNESDAY SESSIONS VOLUME I

Applications of Lexical Link Analysis Web Service for Large-Scale Automation, Validation, Discovery, Visualization, and Real-Time Program-Awareness

Ying Zhao, Shelley Gallup, and Douglas MacKinnon
Naval Postgraduate School

Published April 30, 2012

Report Documentation Page

Form Approved OMB No. 0704-0188

1. REPORT DATE **30 APR 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE **Applications of Lexical Link Analysis Web Service for Large-Scale Automation, Validation, Discovery, Visualization, and Real-Time Program-Awareness**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Monterey,CA,93943**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

DoD acquisition is an extremely complex system, comprised of myriad stakeholders processes, people, activities, and organizational structures. Processes within this complex system are encumbered by the continuous development of large amounts of unstructured and unformatted acquisition program data, which is narrowly useful, but difficult to aggregate across the ?enterprise.? Yet, acquisition analysts and decision-makers must analyze all types and spectrums of the available data to obtain a complete and understandable picture. This is a kind of systems non-congruence that has been difficult to overcome. For those embedded within the complexities of the acquisition community, this can be a daunting, if not impossible task. We will apply a data-driven automation system, namely, Lexical Link Analysis (LLA) to facilitate acquisition researchers and decision-makers to recognize important connections (concepts) that form patterns derived from dynamic, ongoing data collection. The LLA technology and methodology is used to uncover and display relationships among competing programs and Navy-driven requirements. In the past year, we tested our method using samples of acquisition data for visualization and validity. LLA successfully discovered statistically significant correlations, and automatically extracted lexical links, thus improving acquisition professionals? knowledge. This otherwise might have required expensive?and sometimes scarce?manpower to perform (e.g., asking many contractors, continually looking through documentation, and adding excerpts to categories of interest in various spreadsheets). We also developed LLA into a web service this year and have developed use cases for large-scale LLA applications. We report one use case and the status of the web service in this paper.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**95**	

The research presented at the symposium was supported by the acquisition chair of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research or to become a research sponsor, please contact:

NPS Acquisition Research Program
Attn: James B. Greene, RADM, USN, (Ret.)
Acquisition Chair
Graduate School of Business and Public Policy
Naval Postgraduate School
Monterey, CA 93943-5103
Tel: (831) 656-2092
Fax: (831) 656-2253
E-mail: jbgreene@nps.edu

Copies of the Acquisition Research Program's sponsored research reports may be printed from our website (www.acquisitionresearch.net).

Preface & Acknowledgements

Welcome to our Ninth Annual Acquisition Research Symposium! This event is the highlight of the year for the Acquisition Research Program (ARP) here at the Naval Postgraduate School (NPS) because it showcases the findings of recently completed research projects—and that research activity has been prolific! Since the ARP's founding in 2003, over 800 original research reports have been added to the acquisition body of knowledge. We continue to add to that library, located online at www.acquisitionresearch.net, at a rate of roughly 140 reports per year. This activity has engaged researchers at over 60 universities and other institutions, greatly enhancing the diversity of thought brought to bear on the business activities of the DoD.

We generate this level of activity in three ways. First, we solicit research topics from academia and other institutions through an annual Broad Agency Announcement, sponsored by the USD(AT&L). Second, we issue an annual internal call for proposals to seek NPS faculty research supporting the interests of our program sponsors. Finally, we serve as a "broker" to market specific research topics identified by our sponsors to NPS graduate students. This three-pronged approach provides for a rich and broad diversity of scholarly rigor mixed with a good blend of practitioner experience in the field of acquisition. We are grateful to those of you who have contributed to our research program in the past and hope this symposium will spark even more participation.

We encourage you to be active participants at the symposium. Indeed, active participation has been the hallmark of previous symposia. We purposely limit attendance to 350 people to encourage just that. In addition, this forum is unique in its effort to bring scholars and practitioners together around acquisition research that is both relevant in application and rigorous in method. Seldom will you get the opportunity to interact with so many top DoD acquisition officials and acquisition researchers. We encourage dialogue both in the formal panel sessions and in the many opportunities we make available at meals, breaks, and the day-ending socials. Many of our researchers use these occasions to establish new teaming arrangements for future research work. In the words of one senior government official, "I would not miss this symposium for the world as it is the best forum I've found for catching up on acquisition issues and learning from the great presenters."

We expect affordability to be a major focus at this year's event. It is a central tenet of the DoD's Better Buying Power initiatives, and budget projections indicate it will continue to be important as the nation works its way out of the recession. This suggests that research with a focus on affordability will be of great interest to the DoD leadership in the year to come. Whether you're a practitioner or scholar, we invite you to participate in that research.

We gratefully acknowledge the ongoing support and leadership of our sponsors, whose foresight and vision have assured the continuing success of the ARP:

- Office of the Under Secretary of Defense (Acquisition, Technology, & Logistics)
- Director, Acquisition Career Management, ASN (RD&A)
- Program Executive Officer, SHIPS
- Commander, Naval Sea Systems Command
- Program Executive Officer, Integrated Warfare Systems
- Army Contracting Command, U.S. Army Materiel Command
- Office of the Assistant Secretary of the Air Force (Acquisition)

- Office of the Assistant Secretary of the Army (Acquisition, Logistics, & Technology)
- Deputy Director, Acquisition Career Management, U.S. Army
- Office of Procurement and Assistance Management Headquarters, Department of Energy
- Director, Defense Security Cooperation Agency
- Deputy Assistant Secretary of the Navy, Research, Development, Test & Evaluation
- Program Executive Officer, Tactical Aircraft
- Director, Office of Small Business Programs, Department of the Navy
- Director, Office of Acquisition Resources and Analysis (ARA)
- Deputy Assistant Secretary of the Navy, Acquisition & Procurement
- Director of Open Architecture, DASN (RDT&E)
- Program Executive Officer, Littoral Combat Ships

We also thank the Naval Postgraduate School Foundation and acknowledge its generous contributions in support of this symposium.

James B. Greene Jr.
Rear Admiral, U.S. Navy (Ret.)

Keith F. Snider, PhD
Associate Professor

Panel 7. Predicting Performance and Interdependencies in Complex Systems Development

Wednesday, May 16, 2012	
1:45 p.m. – 3:15 p.m.	**Chair: Mark Kryzsko,** Deputy Director, Enterprise Information and Office of the Secretary of Defense Studies, Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics ***Facilitating Decision Choices With Cascading Consequences in Interdependent Networks*** Anita Raja, Mohammad Rashedul Hasan, and Mary Maureen Brown *University of North Carolina at Charlotte* ***Applications of Lexical Link Analysis Web Service for Large-Scale Automation, Validation, Discovery, Visualization, and Real-Time Program-Awareness*** Ying Zhao, Shelley Gallup, Douglas MacKinnon *Naval Postgraduate School* ***Acquisition Management for System-of-Systems: Requirement Evolution and Acquisition Strategy Planning*** Seung Yeob Han, Zhemei Fang, and Daniel DeLaurentis *Purdue University*

Mark Kryzsko—Mr. Krzysko serves as the deputy director of the Enterprise Information and Office of the Secretary of Defense Studies. In this senior leadership position, he oversees Federally Funded Research and Development Centers and directs data governance, technical transformation, and shared services efforts to make timely, authoritative acquisition information available to support oversight of the Department of Defense's major programs—a portfolio totaling more than $1.6 trillion of investment funds over the life cycle of the programs.

Preceding his current position, Mr. Krzysko served as ADUSD for business transformation, providing strategic guidance for re-engineering the Department's business system investment decision-making processes. He also served as ADUSD for strategic sourcing & acquisition processes and as director of the Supply Chain Systems Transformation Directorate, championing and facilitating innovative uses of information technologies to improve and streamline the supply chain process for the Department of Defense. As the focal point for supply chain systems, he was responsible for transformation, implementation, and oversight of enterprise capabilities for the acquisition, logistics, and procurement communities. In addition, Mr. Krzysko served as advisor to the deputy under secretary of defense for business transformation on supply chain matters and as the functional process proponent to the Department's business transformation efforts, resulting in the establishment of the Business Transformation Agency.

In March 2002, Mr. Krzysko joined the Defense Procurement and Acquisition Policy office as deputy director of e-business. As the focal point for the acquisition domain, he was responsible for oversight and transformation of the acquisition community into a strategic business enterprise. This included driving the adoption of e-business practices across the Department, leading the move to modernize processes and systems, and managing the investment review process and portfolio of business systems. Mr. Krzysko served as the division director of Electronic Commerce Solutions for the Naval Air Systems Command from June 2000 to March 2002. From April 1991 until March 2000, Mr. Krzysko served in various senior-level acquisition positions at the Naval Air Systems Command, including contracting officer of F/A-18 foreign military sales, F/A-18 developmental programs, and the F-14. In addition, he served as program manager of Partnering, the Acquisition Business Process Re-engineering Effort, and as acquisition program manager for the Program Executive Office for Tactical Aircraft.

Mr. Krzysko began his career in the private sector in various executive and managerial positions, including assistant managing director for Lord & Taylor Department Stores and operations administrator for Woodward & Lothrop Department Stores. Mr. Krzysko holds a Bachelor of Science degree in finance from the University of Maryland University College, College Park, MD, and a Master of General Administration degree in financial management from the same institution.

Applications of Lexical Link Analysis Web Service for Large-Scale Automation, Validation, Discovery, Visualization, and Real-Time Program-Awareness

Ying Zhao—Dr. Zhao is a research associate professor at the Naval Postgraduate School. Dr. Zhao joined NPS in May 2009. Her research is focused on knowledge management approaches such as data text mining using lexical link analysis, search and visualization for system self-awareness, decision-making, and collaboration. She received her PhD in mathematics from MIT and co-founded Quantum Intelligence, Inc. She has been principal investigator (PI) for six DoD Small Business Innovation Research (SBIR) awarded contracts, and is a co-author of two patents in knowledge pattern search from networked agents, fusion, and visualization for multiple anomaly detection systems. [yzhao@nps.edu]

Shelley Gallup—Dr. Gallup is a research associate professor at the Naval Postgraduate School's Department of Information Sciences, and director of Distributed Information and Systems Experimentation (DISE). Dr. Gallup has a multi-disciplinary science, engineering, and analysis background including microbiology, biochemistry, space systems, international relations, strategy and policy, and systems analysis. He returned to academia after retiring from naval service in 1994, and received his PhD in engineering management from Old Dominion University in 1998. Dr. Gallup joined NPS in 1999, bringing his background in systems analysis, naval operations, military systems, and experimental methods first to the Fleet Battle Experiment series (1999–2002), then to the FORCEnet experimentation in the Trident Warrior series of experiments (2003–present). [spgallup@nps.edu]

Douglas MacKinnon—Dr. MacKinnon is a research associate professor at the Naval Postgraduate School (NPS). Dr. MacKinnon led an NPS research team to assess new MDA, spiral-1 technologies being fielded by PEO C4I developing original decision matrix structures and metrics structures to leverage the new technology. He has also led the assessment of TPED (tasking, planning, exploitation, and dissemination) process during field experiments Empire Challenge 2008 and 2009 (EC08/09). He holds a PhD from Stanford University, conducting theoretic and field research in knowledge management (KM). He has served as the program manager for two major government projects of over $50 million each, implementing new technologies while reducing manpower requirements. He has served over 20 years as a naval surface warfare officer, amassing over eight years at sea, serving in four U.S. Navy warships with five major, underway deployments. [djmackin@nps.edu]

Abstract

DoD acquisition is an extremely complex system, comprised of myriad stakeholders, processes, people, activities, and organizational structures. Processes within this complex system are encumbered by the continuous development of large amounts of unstructured and unformatted acquisition program data, which is narrowly useful, but difficult to aggregate across the “enterprise.” Yet, acquisition analysts and decision-makers must analyze all types and spectrums of the available data to obtain a complete and understandable picture. This is a kind of systems *non-congruence* that has been difficult to overcome. For those embedded within the complexities of the acquisition community, this can be a daunting, if not impossible, task. We will apply a data-driven automation system, namely, Lexical Link Analysis (LLA) to facilitate acquisition researchers and decision-makers to recognize important connections (concepts) that form patterns derived from dynamic, ongoing data collection. The LLA technology and methodology is used to uncover and display relationships among competing programs and Navy-driven requirements. In the past year, we tested our method using samples of acquisition data for visualization and validity. LLA successfully discovered statistically significant correlations, and automatically extracted lexical links, thus improving acquisition professionals’ knowledge. This otherwise might have required expensive—and sometimes scarce—manpower to perform (e.g., asking many contractors, continually looking through documentation, and adding excerpts to categories of interest in various

spreadsheets). We also developed LLA into a web service this year and have developed use cases for large-scale LLA applications. We report one use case and the status of the web service in this paper.

Significance of the Research

We have conducted two research projects to date, namely "Towards Real-Time Program-Awareness via Lexical Analysis" (Phase I; Zhao et al., 2010) and "A Web Service Implementation for Large-Scale Automation, Visualization and Real-Time Program-Awareness via Lexical Link Analysis" (Phase II; Zhao et al., 2011). This follow-up research (Phase III) extends the work of the previous two projects.

We have attempted to develop and frame our research efforts in and around research questions in the following categories: conceptual, focused, theory development, and methodology, in the past three years within the Acquisition Research Program. The questions and research results are summarized in the following sections.

Conceptual

- How can the information that emerges from the acquisition process be used to produce overall awareness of the fit between programs, projects, and systems, and of the needs for which they were intended?

Acquisition research has increased in component, organizational, technical, and management complexity. It is difficult for acquisition professionals to remain continuously aware of their decision-making domains because information is overwhelming and dynamic. According to the *Chairman of the Joint Chiefs of Staff Instruction for Joint Capabilities Integration and Development System (JCIDS*; CJCS, 2009), there are three key processes in the DoD that must work in concert to deliver the capabilities required by the warfighters: the requirements process; the acquisition process; and the Planning, Programming, Budget, and Execution (PPBE) process.

Each process produces a large amount of data in an unstructured manner; for example, the warfighters' requirements are documented in Universal Joint Task Lists (UJTLs), Joint Capability Areas (JCAs), and Urgent Need Statements (UNSs). These requirements are processed in the JCIDS to become projects and programs, which should result in products such as weapon systems that meet the warfighters' needs. Program data are stored in the Defense Acquisition System (DAS). Programs are divided into Major DoD Acquisition Programs (MDAP), Acquisition Category II (ACATII), and so forth. Program Elements (PE) are the documents used to fund programs yearly through the congressional budget justification process. The data is too voluminous, too unformatted, and too unstructured to be easily digested and understood—even by a team of acquisition professionals.

In precise terms, we observed that there were three important processes that seem fundamentally disconnected. Specifically, they were the congressional budgeting justification process (such as information contained within the PEs), the acquisition process (such as information in the MDAP and ACATII), and the warfighters' requirements (such as information in UNSs and in UJTLs), as shown in Figure 1. Yet, these were not analyzed and compared together in a dynamic, holistic methodology that could keep up with changes and reflect patterns of relationships.

There had been little previous effort to integrate the data in these three components. For example, the Matrix Mapping Tool (MMT; Dahmann et al., 2005) included MDAP, UJTL, and JCA, yet did not include PE. Furthermore, in MMT, the links among programs and the matches to UJTL were extracted manually and were therefore not updated in a timely fashion. We employed the LLA automation methodology to analyze more data, thereby achieving a better outcome and provided dynamic, real-time integration. We focused our efforts on demonstrating validation and visualization and on providing insights for decision-makers in three areas, as illustrated in Figure 1.

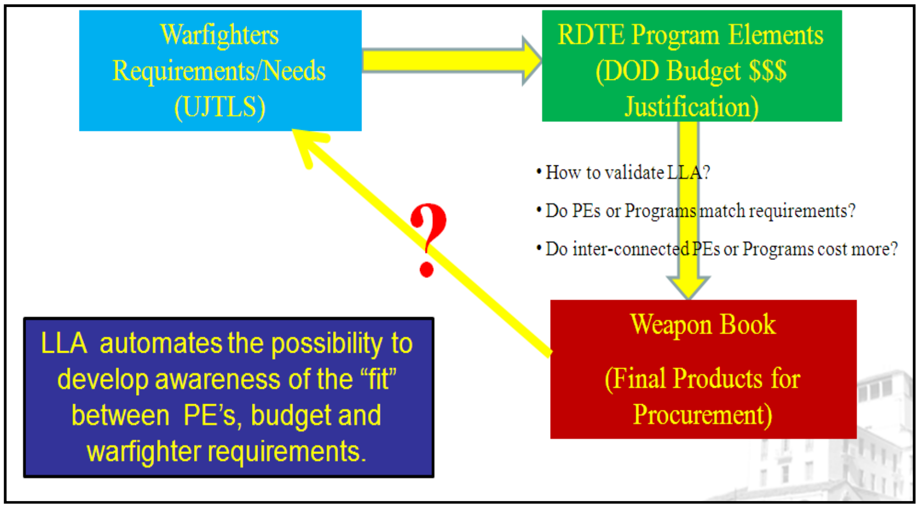


Figure 1. Determining Business Processes Links From Requirements to DoD Budget Justification to Final Products

- If a higher level of awareness is possible, how will that enable system-level regulation of programs, projects, and systems, for improvement of the acquisition system?

To realize the potential of the LLA method, we first established the validity of the method in the context of realistic, large-scale data sets, which include the budgeting process through PEs to the acquisition process via acquisition programs (MDAPs, ACATIIs) to the warfighters' requirements (UNS, UJTL, etc.). We implemented an LLA platform from which to periodically present all the information in a single location so that users can view the trends based on the data in each of the three areas. We gathered the most recent documents in three areas from the following sources:

1. PEs: http://www.dtic.mil/descriptivesum/
2. MDAPs & ACATIIs: http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf, http://www.fas.org/man/dod-101/sys/land/wsh2007/index.html, http://www.acq.osd.mil/ara/am/sar/
3. UJTLs: http://www.dtic.mil/doctrine/jel/cjcsd/cjcsm/m350004d.pdf

Result 1

We found that the Pearson correlation between the links identified by human analysts and by the LLA method was 0.57 with a *p*-value = 10e-7 (Zhao et al., 2010, 2011). LLA was used to correctly predict 80% of the links identified by the human analysts.

High correlation of LLA results with the link analysis done by human analysts makes it possible for automation, saving human power and improving responsiveness. Automation is achieved via computer program or software *agent(s)* to perform LLA frequently—and in near real-time. Agent learning makes it possible to reach real-time; visualization correlates lexical links to core measures; features and patterns are discovered over time for the system as a whole. We can take advantage of the data in motion (Twitter and social media sites) and RSS feed data to build a better picture of real-time program awareness.

Much of text analysis depends on initially searching the available internet. At this point, our efforts are sometimes compared to those of a typical search engine. One of the disadvantages of conventional search engines is that they typically sort documents based on the popularity of documents among linked documents, not based on semantics. Therefore, it does not satisfy complete search needs nor determine relevance if the links among the documents are not available. For example, the content in the forum is not cross-linked, therefore, the discovered or *revealed* topics or themes cannot be found as prioritized results, if conventional search engines are used.

Focused

- Based on the normal evolution of documentation and on the current data-based program information, how can requirements (needs) be connected to system capabilities via automated analysis?
- How can requirements gaps be revealed?

Result 2

We took a detailed look at the RDT&E budget modification practice from 2008 to 2009, observed percentage change for the PE, whose number of LLA links to other PEs was larger than 10, was 14%, compared to 40%, whose number of LLA links to other PEs was fewer than 10. This indicated the current practice tended to reduce the budget for the PEs with more links to other PEs and to increase the budget for the ones with less links, allocating resources to avoid interdependencies and overlapping efforts. However, the numbers of LLA links to the UJTLs were much fewer. The PEs that had at least one LLA match to UJTLs had an average percentage cost increase of 10%, compared to 29% for PEs which had no matches. This indicated a need to consider gaps and the warfighters' requirements as priorities in the RDT&E investment (Zhao et al., 2011a, 2011b).

This demonstrated that our approach "discovers" and displays semantic networks and social networks of programs and PEs. It may also discover blind spots of human analysis that are caused by the overwhelming data for human analysts to go through. These findings can be useful as validation and guidance for implementing the DoD's budget reduction planning. The pattern revealed by LLA creates an opportunity to reduce the overall inefficiency of the cost cutting of linking programs with warfighters' requirements, as opposed to the cost cutting which focuses mainly on the big ticket items such as MDAPs.

Theory Development

- How can a correlation between system interdependency (links/relationships) and development costs be determined and exhibited if found?

Result 3

We used the LLA method to generate semantic networks for the PEs, where two PEs are connected if they are discovered to be using similar lexical terms from the LLA method. As shown in Figure 2, which is laid out by the free energy of the network connections, with the more connected programs in the middle, larger sizes of nodes tend to be on the outside, indicating the correlation between independencies of programs and cost increases. The social network links marked by human analysts, in contrast, do not reveal this pattern.

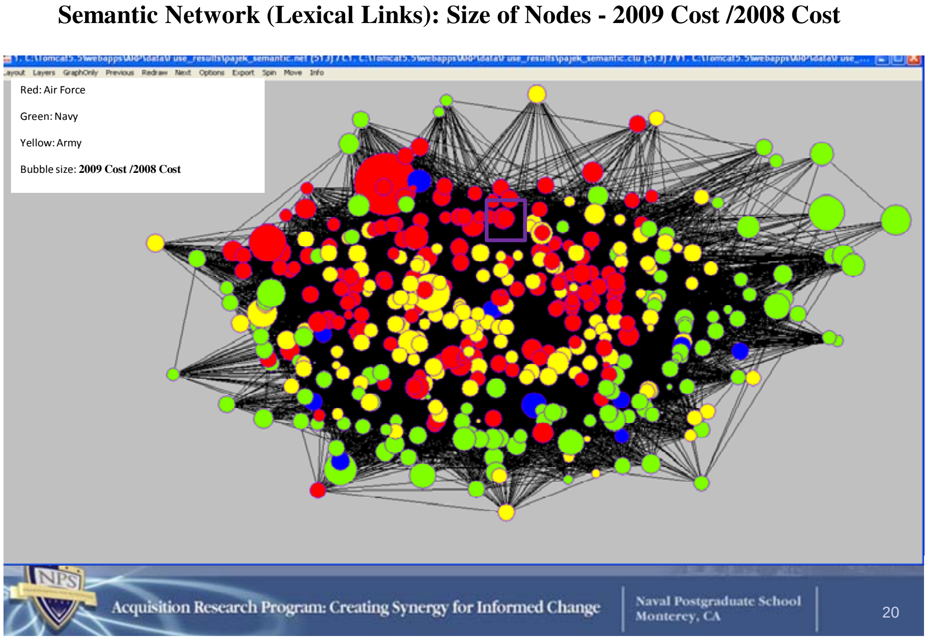


Figure 2. A 3-D View of PEs Identified by the LLA Semantic Network

Methodology (see a full review in the appendix)

- How can we use natural language and other documentation (roughly, unformatted data) to produce visualization of the internal constructs useful for management through Lexical Link Analysis (LLA)?

The LLA method provides the solutions to meet the critical needs of acquisition research. The key advantage is to provide an innovative, near real-time self-awareness system to transfer diversified data services into strategic decision-making knowledge, detailed as follows.

As we continue validating LLA by direct correlation with human analysts' results, we recognize that using LLA to validate human analysis is yet another advantage of our methodology. For instance, LLA may provide different perspectives of links. In the acquisition context, links discovered by human analysts may emphasize component/part connections. They do not necessarily reflect content overlaps; therefore, interdependencies of the programs identified by human analysts (e.g., program managers), might help the programs to stay funded from year to year for the benefit of continuing the program itself, yet may not improve cost reduction for the government. LLA looks for overlapping of the contents in order to improve affordability and meet the requirements of warfighters. Consequently, it provides better results in terms of trust, quality of association, discovery, and breakthrough in the taxonomy of ignorance, organizational boundaries, and organizational reach (Denby & Gammack, 1999).

2012 Phase III Initial Results

The research we have proposed for FY2012 will extend our previous work in the following ways:

1. Build at least two use cases of applications of Lexical Link Analysis Web Service for large-scale automation, validation, discovery, visualization, and real-time program awareness.

2. Demonstrate the methodology for assisting the DoD-wide effort of integrating and maintaining authoritative and accurate acquisition data services in both legacy and new platforms.

Analysis of the Acquisition Research Program (ARP) Data

We started working with the data for the NPS Acquisition Research Program. This is one of the proposed use cases. We have downloaded about 740 publications (from 2003 to 2010) from the website http://www.acquisitionresearch.net.

Each report was labeled manually with a category, for example, "Acquisition Strategy" or "Costing." There are ~160 categories created for this time span (from 2003 to 2010). Figure 3 shows the number of reports in Table 1, using the size of bubbles for each category and year. By observing the bubble chart, we found there are three types of categories:

- steady categories in which the number of reports increased from 2003 to 2010, as shown in Figure 4;
- new and emerging categories in which there were relatively new from 2006 to 2010 compared to 2003 to 2005, as shown Figure 5; and
- die-down categories in which the number of reports reduced from 2006 to 2010 compared to 2003 to 2005, as shown Figure 6.

Table 1. ARP Reports From 2003 to 2010

Year	# of Reports	# of Categories
2003	8	6
2004	27	17
2005	61	34
2006	62	29
2007	143	63
2008	144	68
2009	127	61
2010	184	65

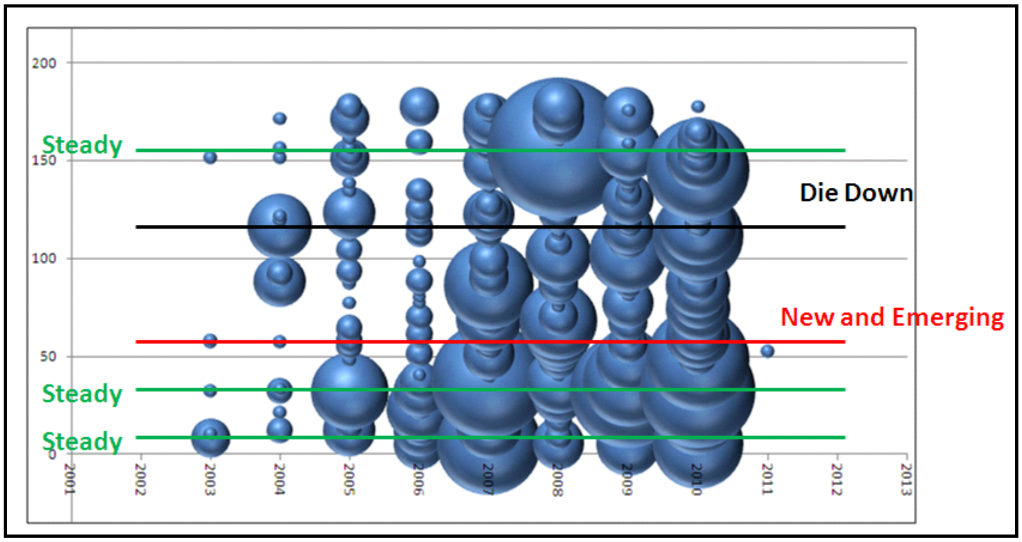


Figure 3. Bubble Chart of the Categories and Time

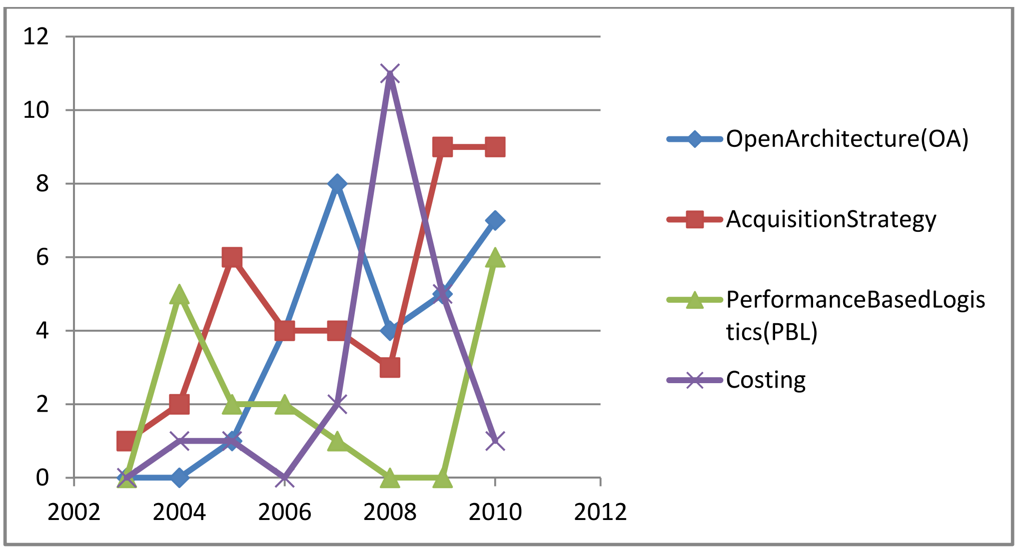


Figure 4. Steady Categories

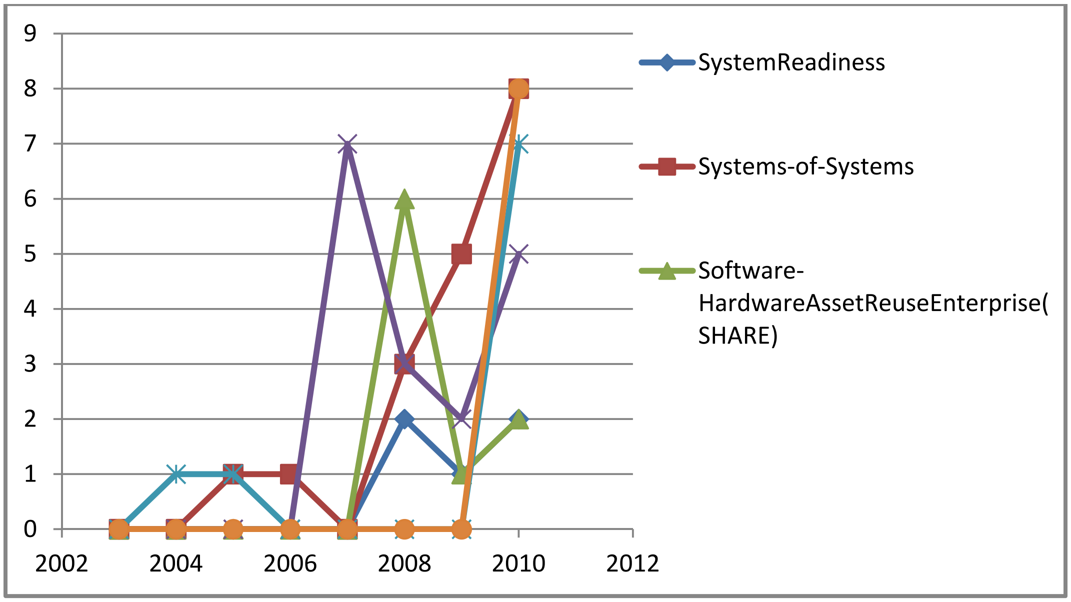


Figure 5. New and Emerging Categories

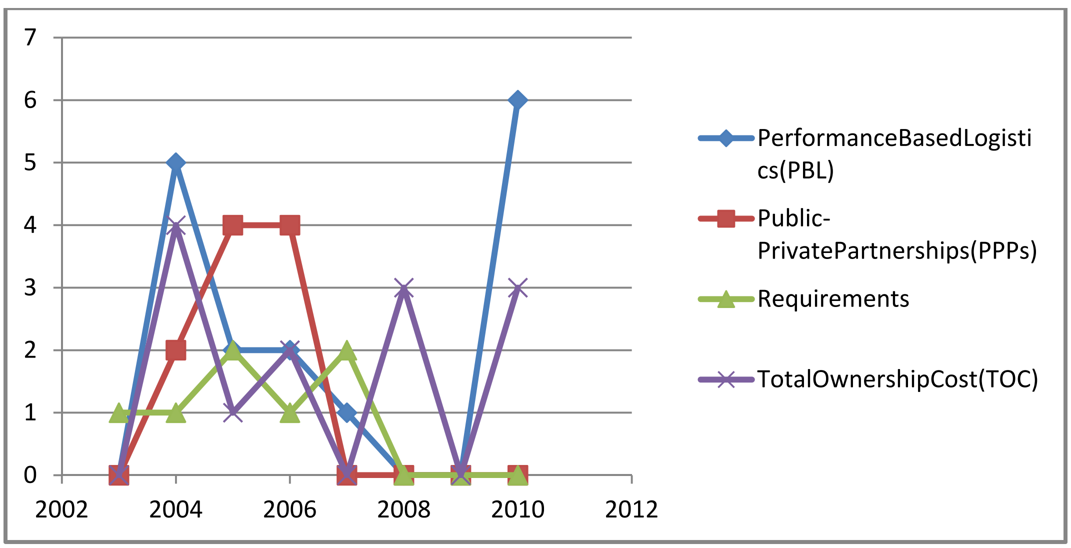


Figure 6. Die-Down Categories

The question is, what are the characteristics for the three types of categories? We used the LLA analysis to examine the changes of lexical links for each category over the years, in an effort to find out the factors that contributed to the dynamics of the three types of categories, especially the characteristics of the steady, new and emerging categories.

To correlate the factors that might be discovered using the LLA method, we used the following methodology and data:

- We first sorted out the existing combinations of year (2003–2009) and 160 categories (e.g., 2003–AcquisitionStrategy and 2004–Outsourcing, etc.). There are a total of 245 such combinations.
- For each combination, we labeled it 1 (*kept*), if the associated category was continued in the following year, e.g., 2003–AcquisitionStrategy is an existing

category, and 2004–AcquisitionStrategy is also one; 0 (*deleted*), if the associated category was not continued in the next year, e.g., 2003–ContractCloseout is an existing category, but 2004–ContractCloseout is not (no reports were classified in the ContractCloseout category in 2004).

The combinations and labels represent the following two decision-makings in the Acquisition Research Program: (1) if a research area or project moves forward from one year to another; and (2) how a research area or project is categorized. By furthering our understanding of how the dynamics of the combinations were *kept* or *deleted* from 2003 to 2010, we hope to shed a light on how the decisions were made in the current process and, more important, to discover the characteristics of research areas, that is, categories that are emerging from the past to the present and to the future.

Previously we introduced, using LLA, how to formulate semantic networks for objects of interest such as PEs. Here we designate the 245 combinations of year-category to be the objects of interest. Figure 7 shows the outputs from LLA showing semantic links between two objects, that is, two year-category combinations with strength calculated from word groups and word hubs.

To simplify the analysis, the links were restricted only within the same year, for example, 2003–AcquistionStrategy is linked to year-categories combination in 2003, not to any other years. We argue here that the simplification is reasonable because the decisions of a categorization and research moving were made heavily based on the information in the current year. Figure 8 shows the semantic networks displayed in the ORA software for 2003 to 2009. Since we want to correlate the *kept* or *delete* labels, we only used the data up to 2010, so no such labels were made for 2010 for this data set. The 2010 data was not used in the semantic network generation and was not included in the 245 combinations.

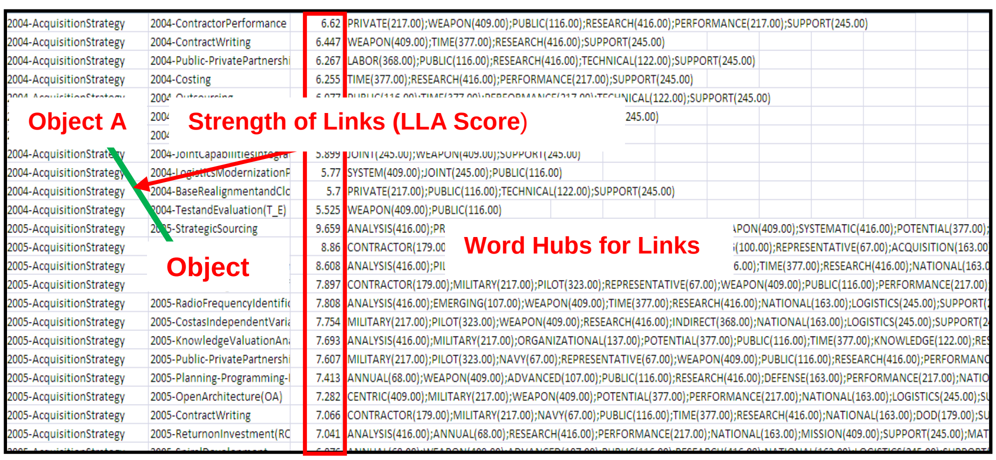


Figure 7. LLA Generates Links, Strength and Associated Word Hubs Between Two Objects (Year-Category Combinations)

As shown in Figure 8, there are seven clusters of semantic networks, representing categories links in the years 2003 to 2009, respectively. Inside each cluster, the strength of the links (LLA scores) between the categories are colored from red to blue. The size of the nodes represents *the total degree of the centrality*, which is the sum of the LLA scores for a given category for that year. The color (red or green) of nodes represents the labels; specifically, it reveals if the category was kept (green) or deleted (red) in the following year.

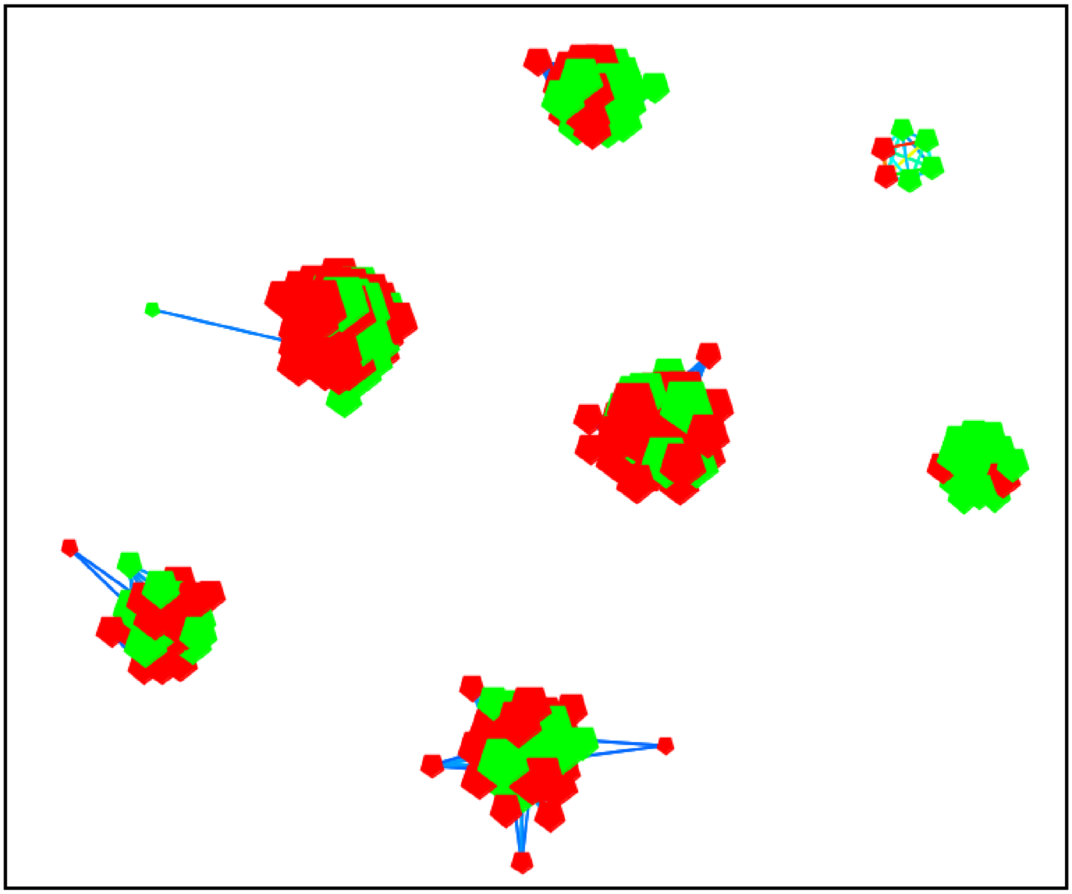


Figure 8. Semantic Networks of Year-Category Combinations From 2003 to 2009

Figure 9 shows a detailed view of the semantic network for 2003 with six categories; two in red were deleted in the following year, and four in green were kept. Initially, looking at each year seems to indicate that the deleted nodes are associated with the "hot" links, that is, links that are in red and orange colors. For example, two such nodes (2003–ContractCloseout and 2003–CostasIndependentVariable) are red for 2003; and one (2004–ContractPerformance) for 2004, as shown in Figure 10. Our hypothesis suggests that emerging categories, the categories that are kept from year to year, might possess the characteristic of having *fewer* overlaps with other categories in any given year. In other words, one of the characteristics of the emerging and steady growing categories might have indicated unique contents compared to the existing information at the time. If a category, represented by the contents in the underlined reports, has too much overlap with other categories at the time, it might be deleted in the following year.

Observing Figure 10, we realized the deleted categories might also be associated with the "cold" links, that is, links that are in green and blue. One such node (2004–LogisticsModenizationProgram) is shown in Figure 10. And this type of node is shown in the border of the graph, indicating that total degree of the node might also be low. Our second hypothesis suggests that categories that are likely to be *deleted* might have *more but weaker* links.

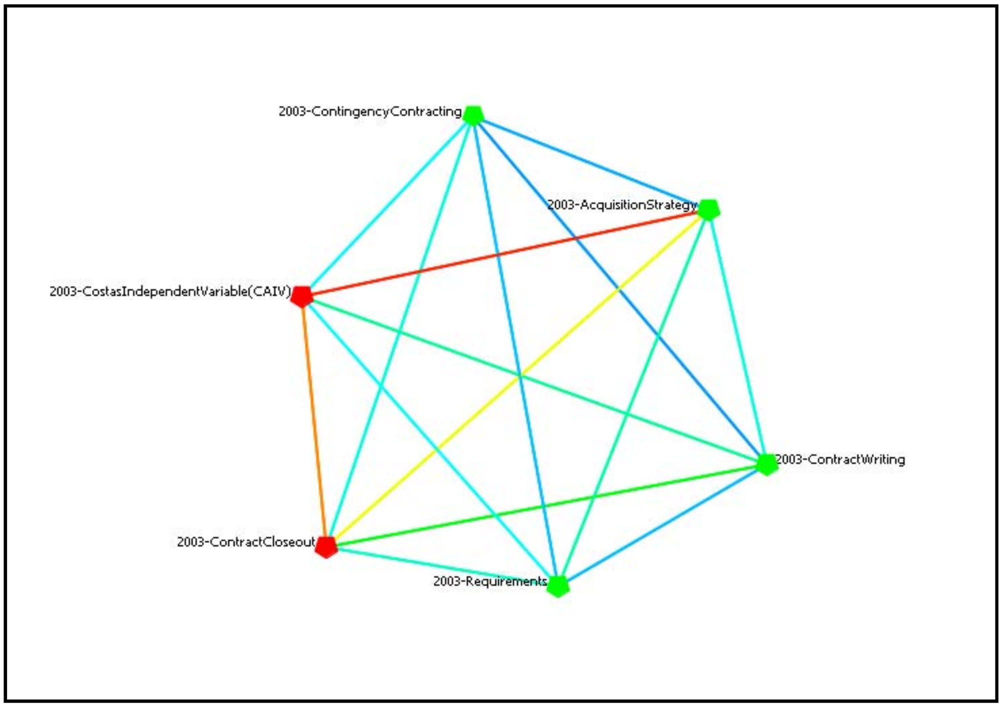


Figure 9. Semantic Network of Year-Category for 2003

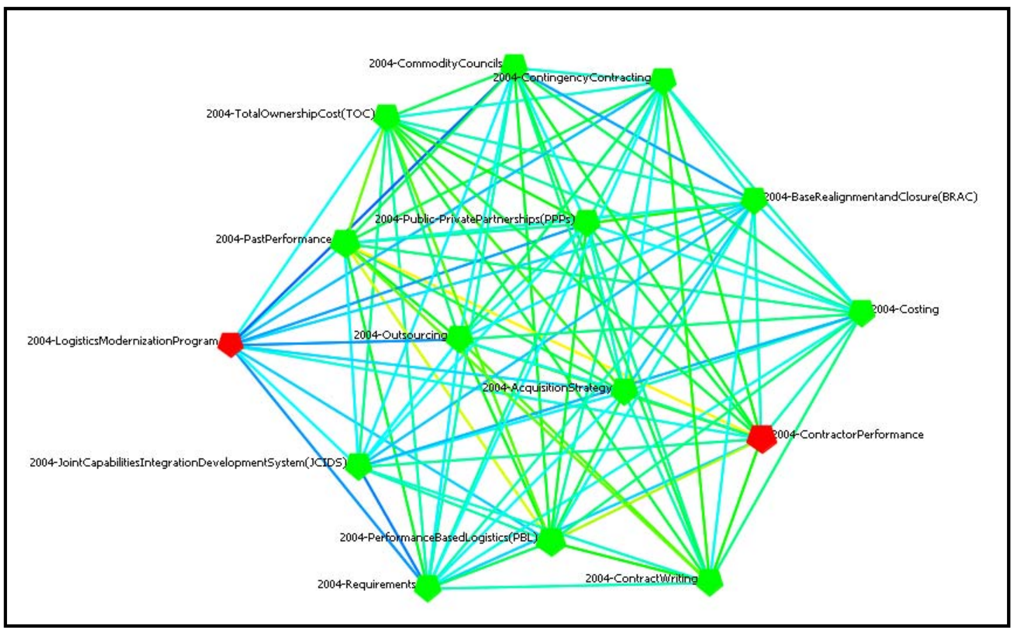


Figure 10. Semantic Network of Year-Category for 2004

We partitioned the data in two ways to validate the hypotheses:

- Divided the 245 nodes into two groups: Group A with 76 nodes, associated only with the links with an LLA score < 7 for the links, and Group B with 169 nodes and only having links with an LLA Score >= 7.
- Computed and sorted the 245 nodes according to the total degree centrality of the network, as shown in Figure 8. Top ranked 76 nodes belong to Group C, and the rest of the nodes belong to Group D.

- Computed the rates (kept/total) of the objects of interest (year-category combinations) that are in each group.

Table 2 shows the summary of the two partitions. As seen here, Group C and Group D have higher kept/total rates than Group A and Group B, respectively. Further statistical tests show that the differences are statistically significant ($p = 0.0017$ and $p = 0.1053$, respectively). This validates our two hypotheses, summarized as follows:

- Categories *kept* (nodes in green) are correlated with at least one hot link with a higher LLA score (threshold set to 7).
- Categories *kept* are correlated with lower total degrees.

In other words, emerging categories tend to form *fewer but stronger* links among the peers. The type of nodes is likely to reside in the so called "Ring of Emergence," as shown in Figure 11 between the red and green circle.

Table 2. Two Ways of the Data Partitions

	Total	**Deleted**	**Kept**	**Kept/Total**	
Group A (LLA Score<7)	76	53	23	0.30	
Group B (LLA Score>=7)	169	84	85	0.50	p=0.0017
Group C (Top Ranked in Total Degree)					
Group D	76	47	29	0.38	
Rest	169	90	79	0.47	p=0.1053

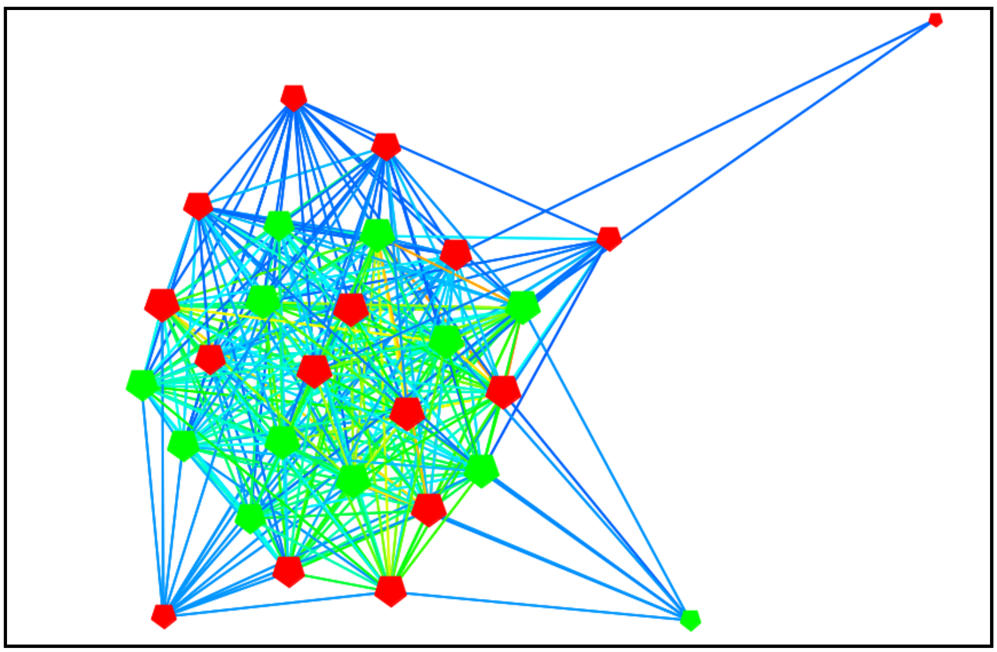


Figure 11. Ring of Emergence

Our future work includes the following:

- discovering the exact conditions to predict the emerging categories by adding other centralities measures of the semantic networks in order to include probability (e.g., rates for Group B and Group D).
- applying automatically discovered themes as categories to see if the same theory applies.

Authoritative and Accurate Acquisition Data Services

In order to integrate with and analyze authoritative and accurate data, we have started to work with the Enterprise Information & Office of the Secretary of Defense (OSD) Studies in the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). The OUSD(AT&L) provides the DoD-wide acquisition community with authoritative and accurate data services. For example, the Defense Technical Information Center (DTIC), Defense Acquisition Management Information Retrieval (DAMIR; http://www.acq.osd.mil/damir/), Acquisition Resources and Analysis (ARA; http://www.acq.osd.mil/ara), and Selected Acquisition Report (SAR; http://www.acq.osd.mil/ara/am/sar) are good sources. *Requirements* data are not included. We packaged the tool and related documentation, as shown in Figure 12.

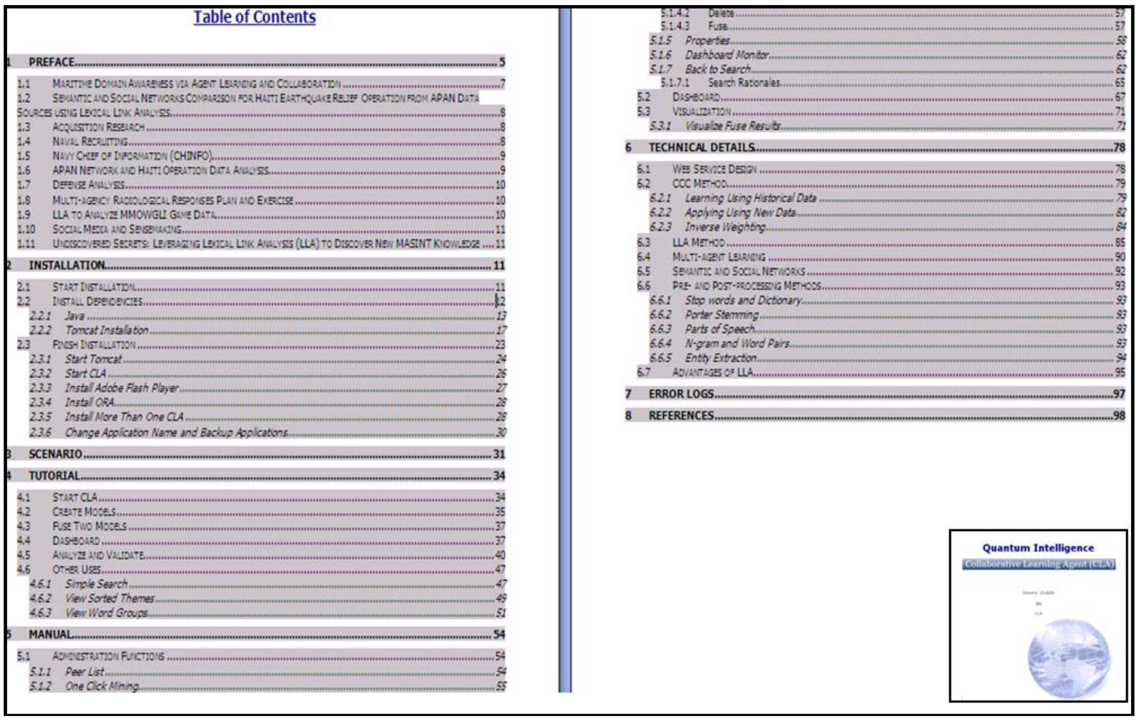


Table of Contents

1 PREFACE 5
1.1 Maritime Domain Awareness via Agent Learning and Collaboration 7
1.2 Semantic and Social Networks Comparison for Haiti Earthquake Relief Operation from APAN Data Sources using Lexical Link Analysis 8
1.3 Acquisition Research 8
1.4 Naval Recruiting 8
1.5 Navy Chief of Information (CHINFO) 9
1.6 APAN Network and Haiti Operation Data Analysis 9
1.7 Defense Analysis 10
1.8 Multi-agency Radiological Responses Plan and Exercise 10
1.9 LLA to Analyze MMOWGLI Game Data 10
1.10 Social Media and Sensemaking 11
1.11 Undiscovered Secrets: Leveraging Lexical Link Analysis (LLA) to Discover New MASINT Knowledge 11
2 INSTALLATION 11
2.1 Start Installation 11
2.2 Install Dependencies 12
2.2.1 Java 13
2.2.2 Tomcat Installation 17
2.3 Finish Installation 23
2.3.1 Start Tomcat 24
2.3.2 Start CLA 26
2.3.3 Install Adobe Flash Player 27
2.3.4 Install ORA 28
2.3.5 Install More Than One CLA 28
2.3.6 Change Application Name and Backup Applications 30
3 SCENARIO 31
4 TUTORIAL 34
4.1 Start CLA 34
4.2 Create Models 35
4.3 Fuse Two Models 37
4.4 Dashboard 37
4.5 Analyze and Validate 40
4.6 Other Uses 47
4.6.1 Simple Search 47
4.6.2 View Sorted Themes 49
4.6.3 View Word Groups 51
5 MANUAL 54
5.1 Administration Functions 54
5.1.1 Peer List 54
5.1.2 One Click Mining 55
5.1.4.2 Delete 57
5.1.4.3 Fuse 57
5.1.5 Properties 58
5.1.6 Dashboard Monitor 62
5.1.7 Back to Search 62
5.1.7.1 Search Rationales 65
5.2 Dashboard 67
5.3 Visualization 71
5.3.1 Visualize Fuse Results 71
6 TECHNICAL DETAILS 78
6.1 Web Service Design 78
6.2 CCC Method 79
6.2.1 Learning Using Historical Data 79
6.2.2 Applying Using New Data 82
6.2.3 Inverse Weighting 84
6.3 LLA Method 85
6.4 Multi-agent Learning 90
6.5 Semantic and Social Networks 92
6.6 Pre- and Post-processing Methods 93
6.6.1 Stop words and Dictionary 93
6.6.2 Porter Stemming 93
6.6.3 Parts of Speech 93
6.6.4 N-gram and Word Pairs 93
6.6.5 Entity Extraction 94
6.7 Advantages of LLA 95
7 ERROR LOGS 97
8 REFERENCES 98

Figure 12. Documentation for the LLA Software

We have also contacted the OSD(AT&L) ARP Enterprise Information and OSD Studies, in the process of evaluating the software and web service FY2012 2nd quarter development circle.

Conclusion

We have summarized the results from Phase I, II, and III for this project. We have focused on showing how to apply the LLA method to the NPS Acquisition Research Program reports from 2003 to 2010. We have discovered the characteristics of emerging categories and validated them with the actual human cognitive data processing and decision-making data. Through new methods of demonstration, we seek to reveal these changes to decision-makers and assist them in making improved decisions in the acquisition process.

References

AutoMap. (2009). AutoMap: Extract, analyze and represent relational data from texts. Retrieved from http://www.casos.cs.cmu.edu/projects/automap/

Blei, D., Ng, A., & Jordan, M. (2003). Latent *Dirichlet* allocation. *Journal of Machine Learning Research, 3*, 993–1022. Retrieved from http://jmlr.csail.mit.edu/papers/volume3/blei03a/blei03a.pdf

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Center for Computational Analysis of Social and Organizational Systems (CASOS). (2009). *AutoMap: Extract, analyze and represent relational data from texts*. Retrieved from http://www.casos.cs.cmu.edu

Chairman of the Joint Chiefs of Staff (CJCS). (2009). *Chairman of the Joint Chiefs of Staff instruction for joint capabilities integration and development system (JCIDS*; J-8 CJCSI 3170.01G). Retrieved from http://www.intelink.sgov.gov/wiki/JCIDS

Dahmann, J., Baldwin, K., Bergin, D., Choudhary, A., Dubon, A., & Eiserman, G. (2005). *Matrix mapping tool (MMT*; White paper). Washington, DC: OUSD(AT&L)/Defense Systems.

Denby, E., & Gammack, J. (1999). *Modelling ignorance levels in knowledge-based decision support*. Retrieved from http://wawisr01.uwa.edu.au/1999/DenbyGammack.pdf

DoD. (2007, February). *Program acquisition costs by weapon system*. Retrieved from http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve information retrieval. In *Proceedings of CHI'88: Conference on Human Factors in Computing* (pp. 281–285). New York, NY: Association for Computing Machinery.

Foltz, P. W. (2002). Quantitative cognitive models of text and discourse processing. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *The handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum.

Gallup, S. P., MacKinnon, D. J., Zhao, Y., Robey, J., & Odell, C. (2009, October 6–8). Facilitating decision making, re-use and collaboration: A knowledge management approach for system self-awareness. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management (IC3K)*. Madeira, Portugal.

Gerber, C. (2005). Smart searching, new technology is helping defense intelligence analysts sort through huge volumes of data. *Military Information Technology*, *9*(9). Retrieved from http://www.mkbergman.com

Girvan, M., & Newman, M. E. J. (2002, June). Community structure in social and biological networks. In *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821–7826.

Lexical analysis. (2010). In *Wikipedia*. Retrieved September 2010 from http://en.wikipedia.org/wiki/Lexical_analysis

Quantum Intelligence (QI). (2009). Collaborative learning agents (CLA). Retrieved from http://www.quantumii.com/qi/cla.html

Vassar. (2010). The significance of the difference between two independent proportions. Retrieved from http://faculty.vassar.edu/lowry/VassarStats.html

Zhao, Y., Gallup, S., & MacKinnon, D. (2010a). Towards real-time program awareness via lexical link analysis. In *Proceedings of the Seventh Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2010b). *Towards real-time program awareness via lexical link analysis* (NPS-AM-10-174). Monterey, CA: Naval Postgraduate School.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011a, May). A web service implementation for large-scale automation, visualization and real-time program-awareness via lexical link analysis. In *Proceedings of the Eighth Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011b). A web service implementation for large-scale automation, visualization and real-time program-awareness via lexical link analysis (NPS-AM-11-186). Monterey, CA: Naval Postgraduate School.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011c, September). System self-awareness and related methods for improving the use and understanding of data within DoD. *Software Quality Professional, 13*(4), 19–31. Retrieved from http://asq.org/pub/sqp/

Appendix: Overview of Lexical Link Analysis

As in military operations, where the term *situational awareness* was coined, we note that our efforts can inform *awareness* of analyzed data in a unique way that helps improve decision-makers' understanding or awareness of its content. We therefore define *awareness* as the cognitive interface between decision-makers and a complex system, expressed in a range of terms or *features*, or specific vocabulary or *lexicon*, to describe the attributes and surrounding environment of the system. Specifically, LLA is a form of text mining in which word meanings represented in lexical terms (e.g., word pairs) can be represented as if they are in a community of a word network.

Link analysis "discovers" and displays a network of word pairs. These word pair networks are characterized by one-, two-, or three-word themes. The weight of each theme is determined by its frequency of occurrence. Figure 13 shows a visualization of lexical links for Systems 1 and 2 of two systems, which are shown in the red box. Unlinked, outer vectors (outside the red box) indicate unique system features. For example, Figure 14 shows that the information from three categories can be compared, and Figure 15 shows that the information from two time periods can be compared.

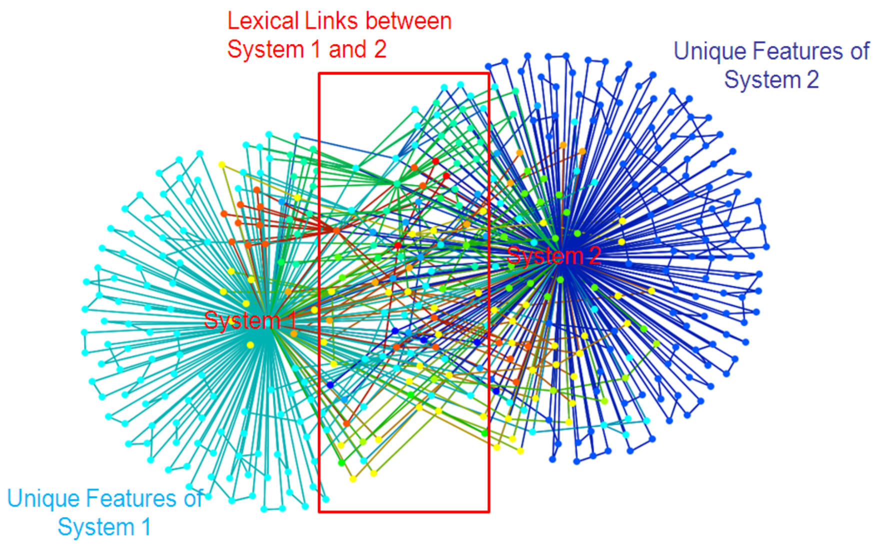


Figure 13. Comparing Two Systems Using LLA

The closeness of the systems in comparison can be visually examined or examined using the Quadratic Assignment Procedure (QAP; Hubert & Schultz, 1976, e.g., in UCINET, Borgatti et al., 2002) to compute the correlation and analyze the structural differences in the two systems, as shown in Figure 16.

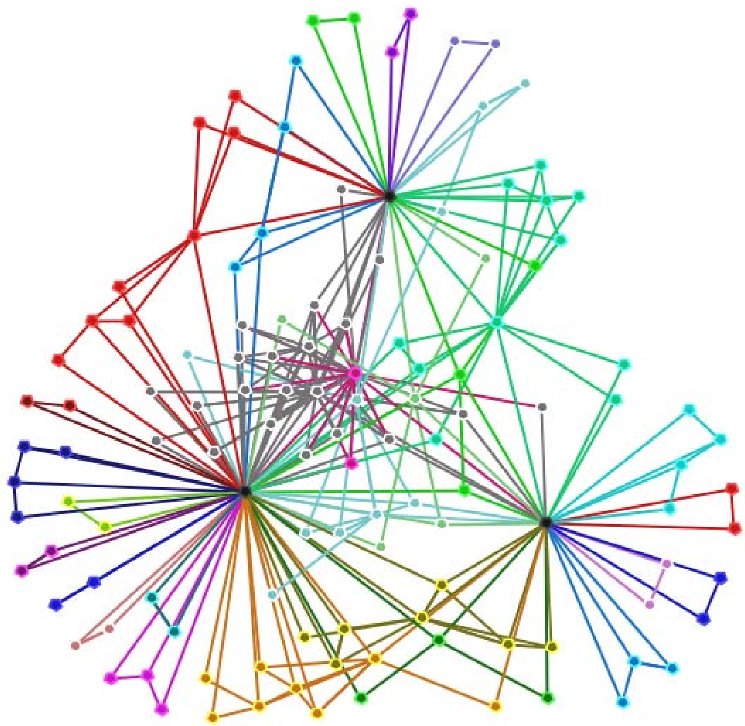


Figure 14. Comparing Three Categories Using LLA

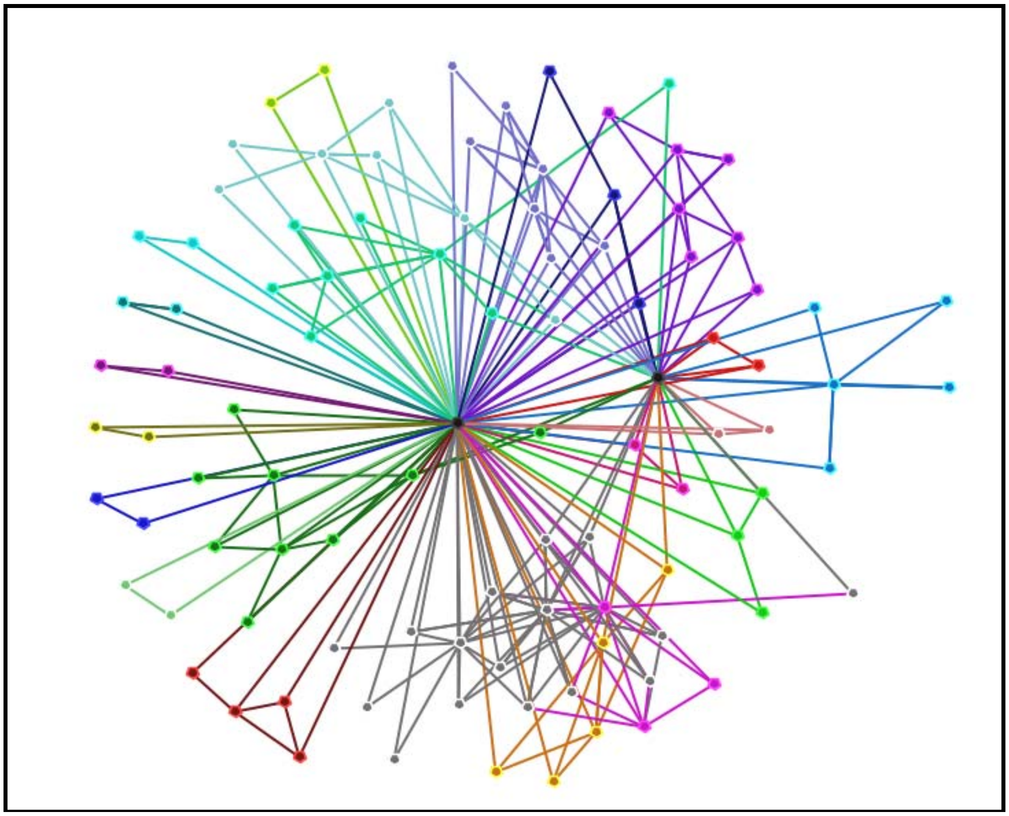


Figure 15. Comparing Two Time Periods

QAP Correlations

		1 lla_n	2 lla_n	3 lla_n	4 lla_n	5 lla_n	6 lla_n	7 lla_n	8 lla_n
1	lla_network_1_2010-AcquisitionStrategy	1.000	0.174	0.156	0.155	0.036	0.111	0.020	0.062
2	lla_network_1_2003-AcquisitionStrategy	0.174	1.000	0.447	0.149	0.052	0.119	0.043	0.089
3	lla_network_1_2004-AcquisitionStrategy	0.156	0.447	1.000	0.111	0.047	0.119	0.051	0.080
4	lla_network_1_2005-AcquisitionStrategy	0.155	0.149	0.111	1.000	0.156	0.084	0.034	0.088
5	lla_network_1_2006-AcquisitionStrategy	0.036	0.052	0.047	0.156	1.000	0.067	0.036	0.056
6	lla_network_1_2007-AcquisitionStrategy	0.111	0.119	0.119	0.084	0.067	1.000	0.097	0.123
7	lla_network_1_2008-AcquisitionStrategy	0.020	0.043	0.051	0.034	0.036	0.097	1.000	0.286
8	lla_network_1_2009-AcquisitionStrategy	0.062	0.089	0.080	0.088	0.056	0.123	0.286	1.000

QAP P-Values

		1 lla_n	2 lla_n	3 lla_n	4 lla_n	5 lla_n	6 lla_n	7 lla_n	8 lla_n
1	lla_network_1_2010-AcquisitionStrategy	0.000	0.020	0.020	0.020	0.020	0.020	0.020	0.020
2	lla_network_1_2003-AcquisitionStrategy	0.020	0.000	0.020	0.020	0.020	0.020	0.020	0.020
3	lla_network_1_2004-AcquisitionStrategy	0.020	0.020	0.000	0.020	0.020	0.020	0.020	0.020
4	lla_network_1_2005-AcquisitionStrategy	0.020	0.020	0.020	0.000	0.020	0.020	0.020	0.020
5	lla_network_1_2006-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.000	0.020	0.020	0.020
6	lla_network_1_2007-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.000	0.020	0.020
7	lla_network_1_2008-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.020	0.000	0.020
8	lla_network_1_2009-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.000

QAP statistics saved as datafile QAP Correlation Results

Figure 16. QAP Correlation via UCINET

Each node, or word hub, represents a system *feature,* and each color refers to the collection of lexicon (features) that describes a concept or theme. The overlapping area nodes are *lexical links*. What is unique here is that LLA constructs these linkages via intelligent agent technology using social network grouping methods.

Figure 17 shows a visualization of LLA with connected keywords or concepts as groups or *themes*. Words are linked as word pairs that appear next to each other in the original documents. Different colors indicate different clusters of word groups. They were produced using a link analysis method—a social network grouping method (Girvan et al., 2001) where words are connected, as shown in a single color, as if they are in a social community. A “hub” is formed around a word centered or connected with a list of other words (“fan-out” words) centered on other hub words. For instance, Figure 18 shows a detailed view of a theme or word group in Figure 17: the words “analysis, research, approach” are connected and centered around other related words. We use three words such as “analysis, research, approach” to label a group.

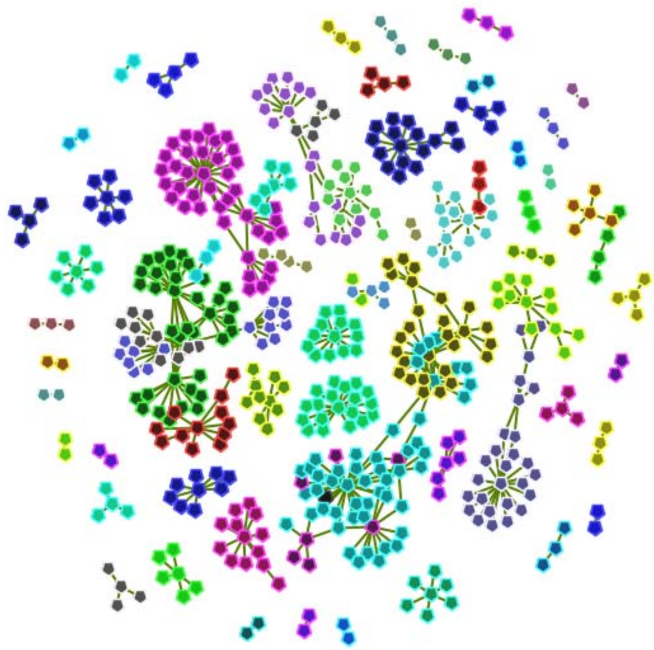


Figure 17. Word and Term of Themes Discovered and Shown in Colored Groups

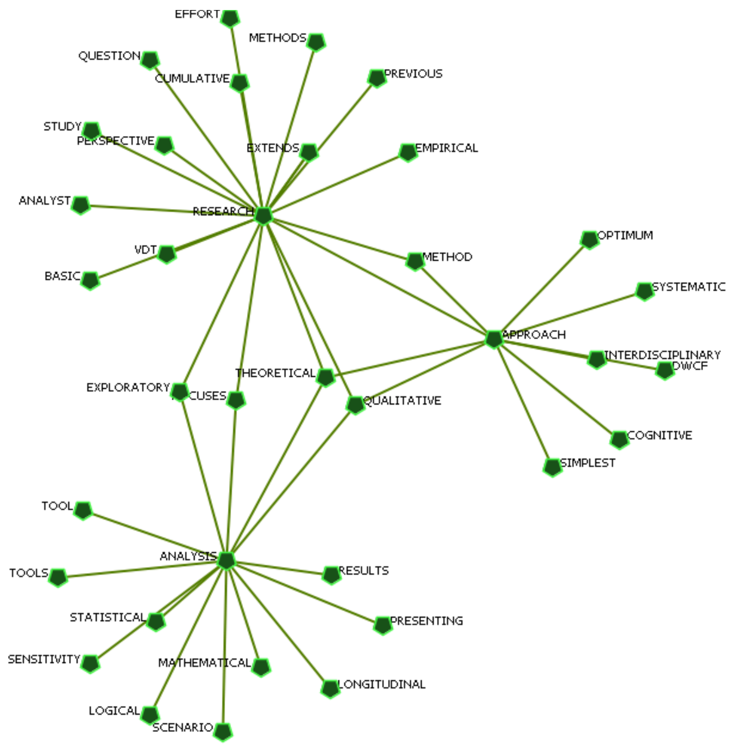


Figure 18. A Detailed View of a Theme or Word Group in Figure 17

The detailed steps of LLA processing include applying collaborative learning agents (CLA) and generating visualizations, including a lexical network visualization via AutoMap (2009), radar visualization, and matrix visualization (Zhao et al., 2010). The following are the steps for performing an LLA:

- Read each set of documents.
- Select feature-like word pairs.
- Apply a social network community finding algorithm (e.g., Newman grouping method; Girvan et al., 2001) to group the word pairs into themes. A theme includes a collection of lexical word pairs connected to each other.
- Compute a "weight" for a theme for the information of a time period, that is, how many word pairs belong to a theme for that time period and for all the time periods.
- Sort theme weights by time, and study the distributions of the themes by time.

General questions that LLA usually answers are as follows:

- Discover themes and topics in the unstructured documents and sort the importance of the themes.
- Discover social and semantic networks of organizations that were involved, and compare the two networks to obtain insights to answer the following questions:
    - What were the organizations involved in the *important* themes?
    - How do semantic networks suggest more potential collaboration when compared to social networks?

In the past year, we began at the Naval Postgraduate School (NPS) by using Collaborative Learning Agents (CLA; QI, 2009) and expanded to other tools, including

AutoMap (CASOS, 2009) for improved visualizations. Results from these efforts arose from leveraging intelligent agent technology via an educational license with Quantum Intelligence, Inc. CLA is a computer-based learning agent, or agent collaboration, capable of ingesting and processing data sources.

The LLA approach is more properly related to Latent Semantic Analysis (LSA; Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988) and Probabilistic Latent Semantic Analysis (PLSA). In the LSA approach, a term-document matrix is the starting point for analysis. The elements of the term-document or feature-object (term as feature, and document as object) matrix are the occurrences of each word in a particular document, that is, $A = [a_{ij}]$, where a_{ij} denotes the frequency in which term *j* occurs in document *i.* The term-document matrix is usually sparse. LSA uses singular value decomposition (SVD) to reduce the dimensionality of the term-document matrix. SVD cannot be applied to the cases in which the vocabulary (the unique number of terms) in the document collection is large. LSA has been widely used to improve information indexing, search/retrieval, and text categorization.

A recent development related to this method is called Latent *Dirichlet* allocation (LDA; Blei, Ng, & Jordan, 2003), which is a generative probabilistic model of a corpus. In LDA, a document is considered to be composed of a collection of words—a “bag of words,” where word order and grammar are not considered important. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a statistical distribution (Dirichlet distribution) over the corpus. Our theme generation from LLA is different than LDA, in which a collection of lexical terms are connected to each other semantically, as if they are in a social community, and social network grouping methods are used to group the words. Our method is easily scaled to analyze a large vocabulary and is generalizable to any sequential data.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

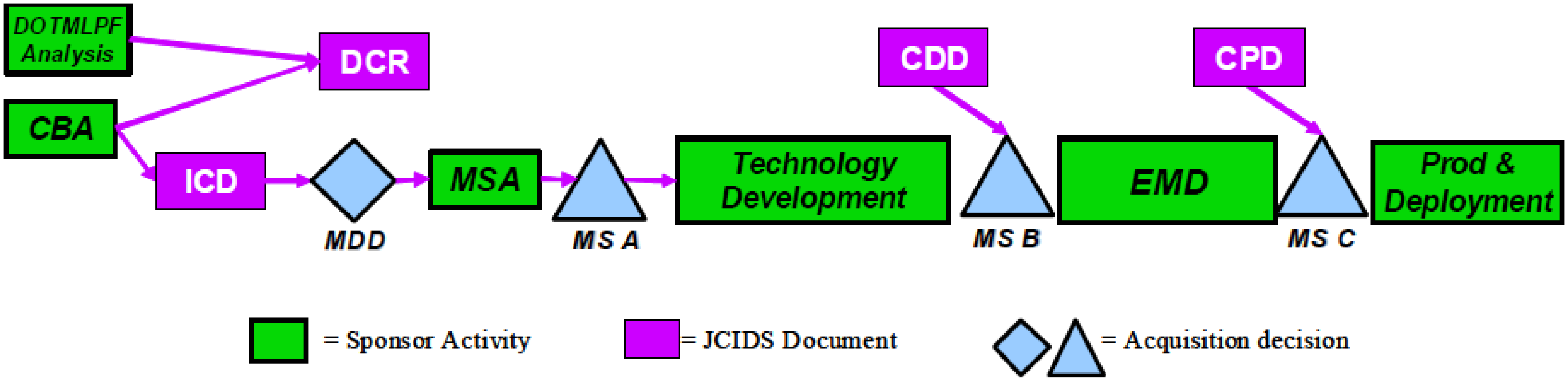
www.acquisitionresearch.net

APPLICATIONS OF LEXICAL LINK ANALYSIS WEB SERVICE FOR LARGE-SCALE AUTOMATION, VALIDATION, DISCOVERY, VISUALIZATION AND REAL-TIME PROGRAM-AWARENESS

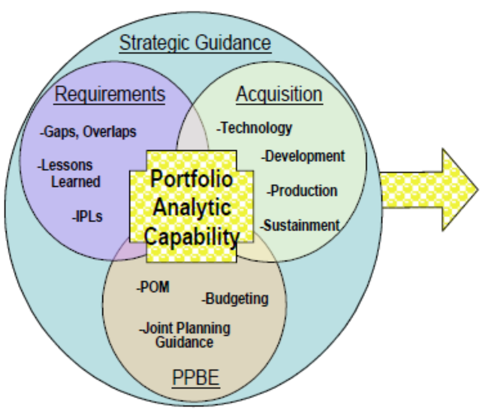
May 16-17, 2012

Dr. Ying Zhao, Dr. Douglas J. MacKinnon, Dr. Shelley P. Gallup
Research Associate Professors
Distributed Information Systems Experimentation, Naval Postgraduate School

Critical Needs: Automation, Validation and Discovery



JCIDS Process and Acquisition Decisions
(J-8 CJCSI 3170.01G)(JCIDS, 2009)



Multiple Portfolio Views:
- Systems vs. Capabilities
- Investment vs. Capabilities
- System Context
- Highly dependent programs (Joint Enablers)
- Procurement Optimization
- S&T vs. future needs
- Sustainment Efficiency
- Market Value

- Data are too voluminous, unformatted and unstructured!
- Need to leverage automation
  - Extract relations among PE, MDAP, and ACATII
  - Extract costs

Research Question

How can the information that emerges from the acquisition process be used to produce overall *awareness* of the *fit* between programs/projects/systems and verify *needs* for which they were intended?

LLA Methodology Can Help!

Warfighters Requirements/Needs (UJTLS)

RDTE Program Elements (DOD Budget $$$ Justification)

?

- How to validate LLA?
- Do PEs or Programs match requirements?
- Do inter-connected PEs or Programs cost more?

LLA automates the possibility to develop awareness of the “fit” between PE’s, budget and warfighter requirements.

Weapon Book

(Final Products for Procurement)

METHODS

System Self-Awareness (SSA)

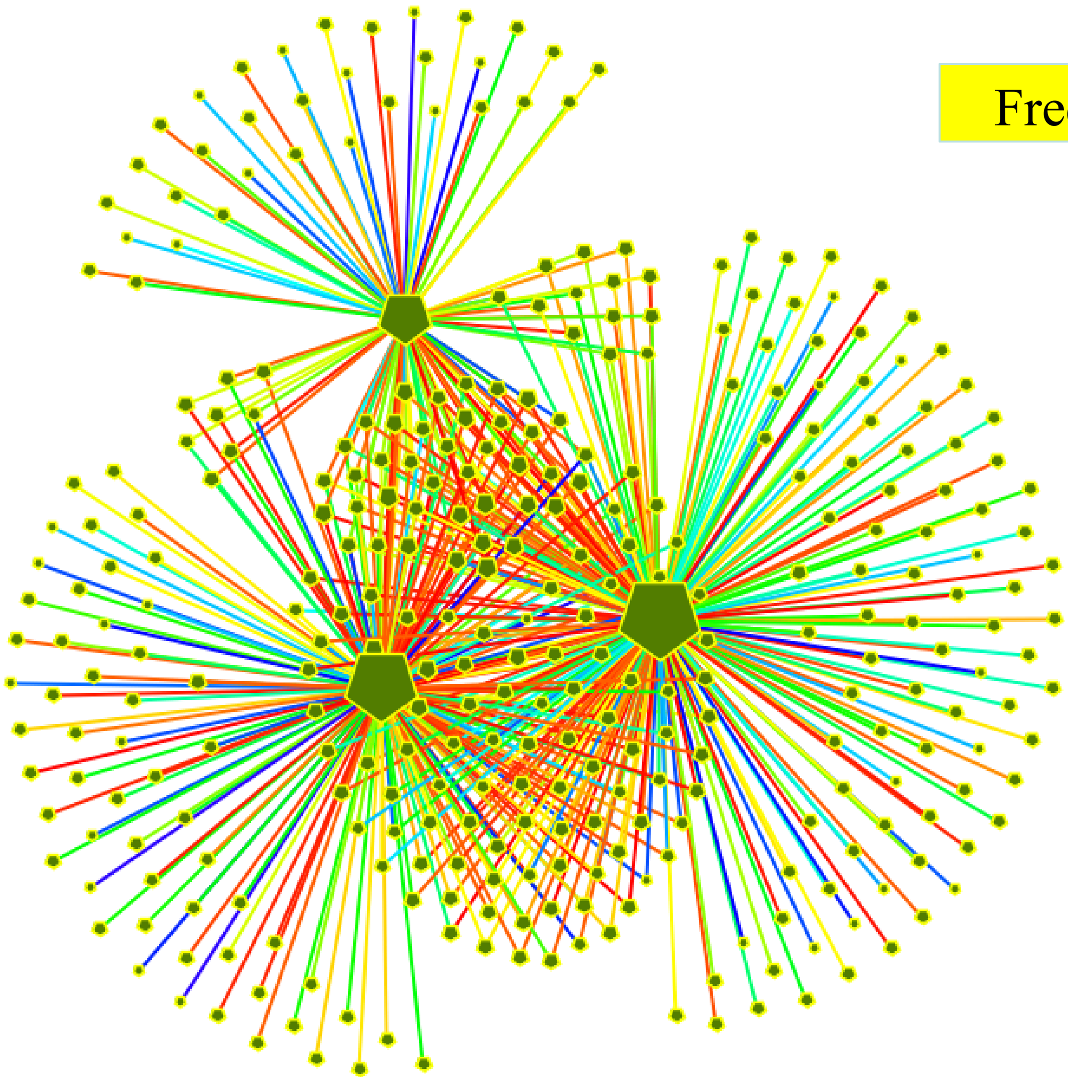
- *Awareness*
  - The cognitive interface between decision makers and a complex system, expressed in a range of terms or “features,” or specific vocabulary or “lexicon,” to describe the attributes and surrounding environment of the system.
- System Self-awareness
  - Complex system’s ability to assess itself within a global context
  - Examples
    - Authority
    - Expertise

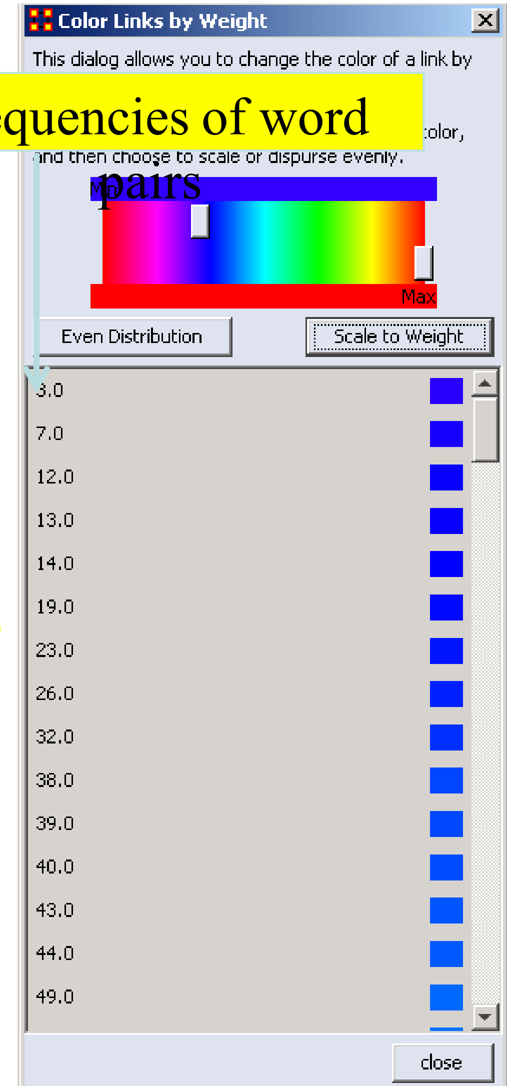
Text Analysis

There are three methods

- Linguistics based methods
  - InXight
- Statistical co-occurrence
- Representation
  - Bag-of-Words (BOW)
  - Text-as-Network (TAN)

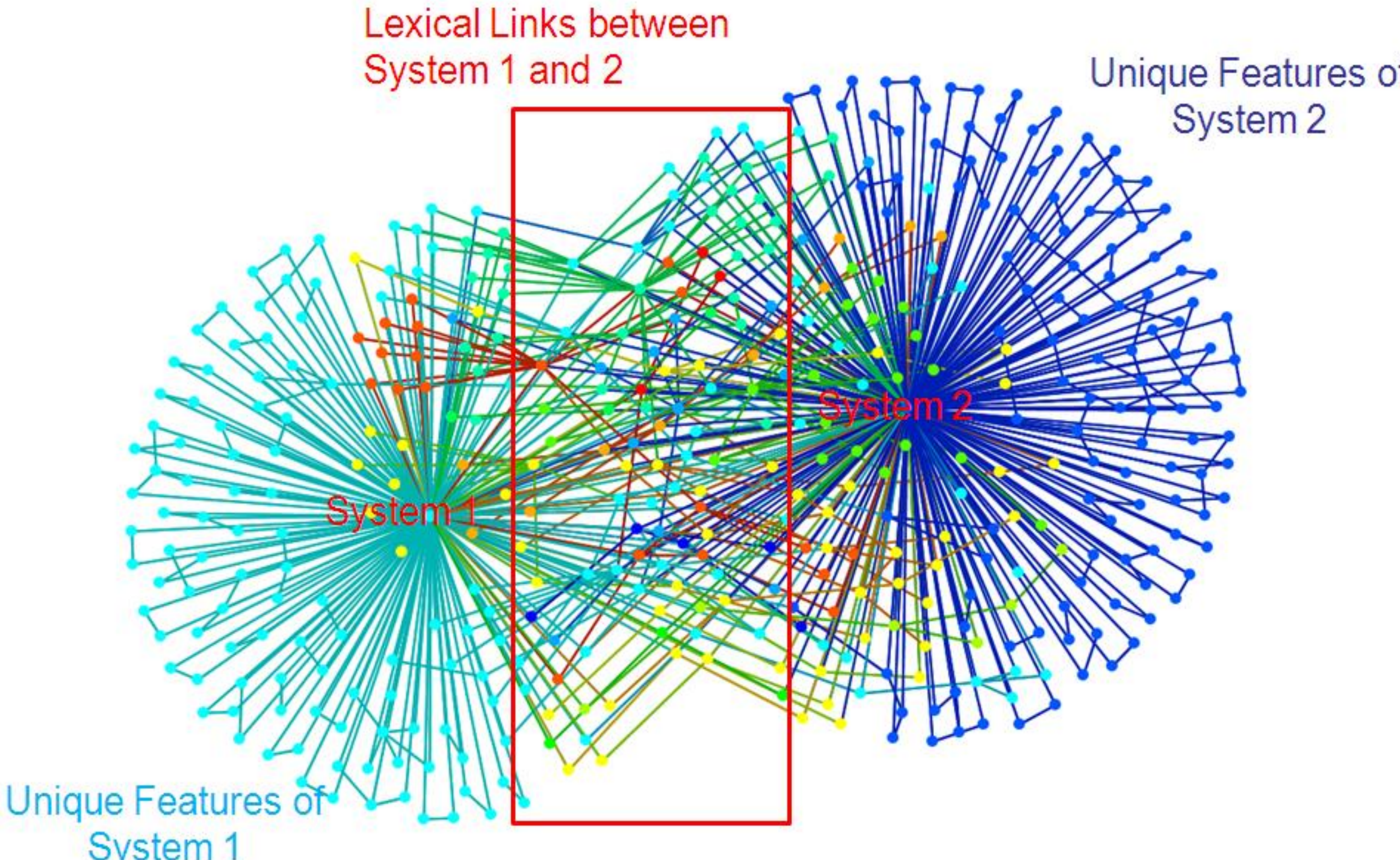
LLA: Bi-gram co-occurrence word pair networks



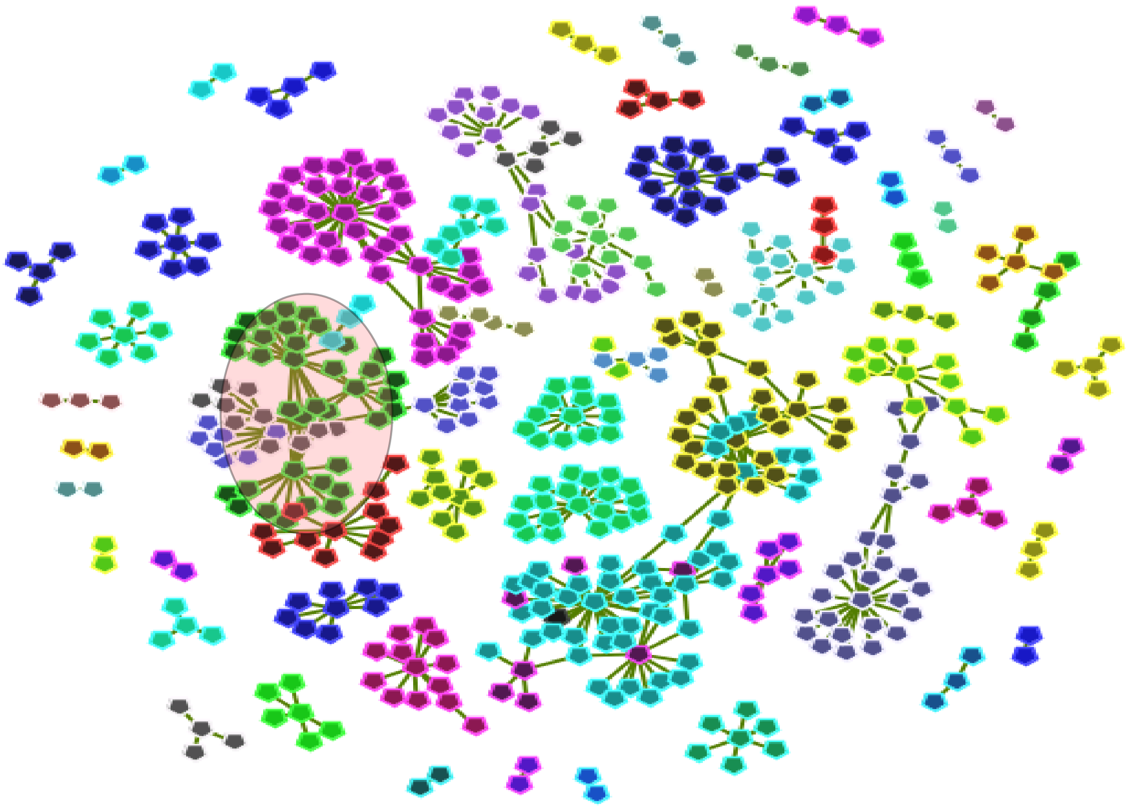


Frequencies of word pairs

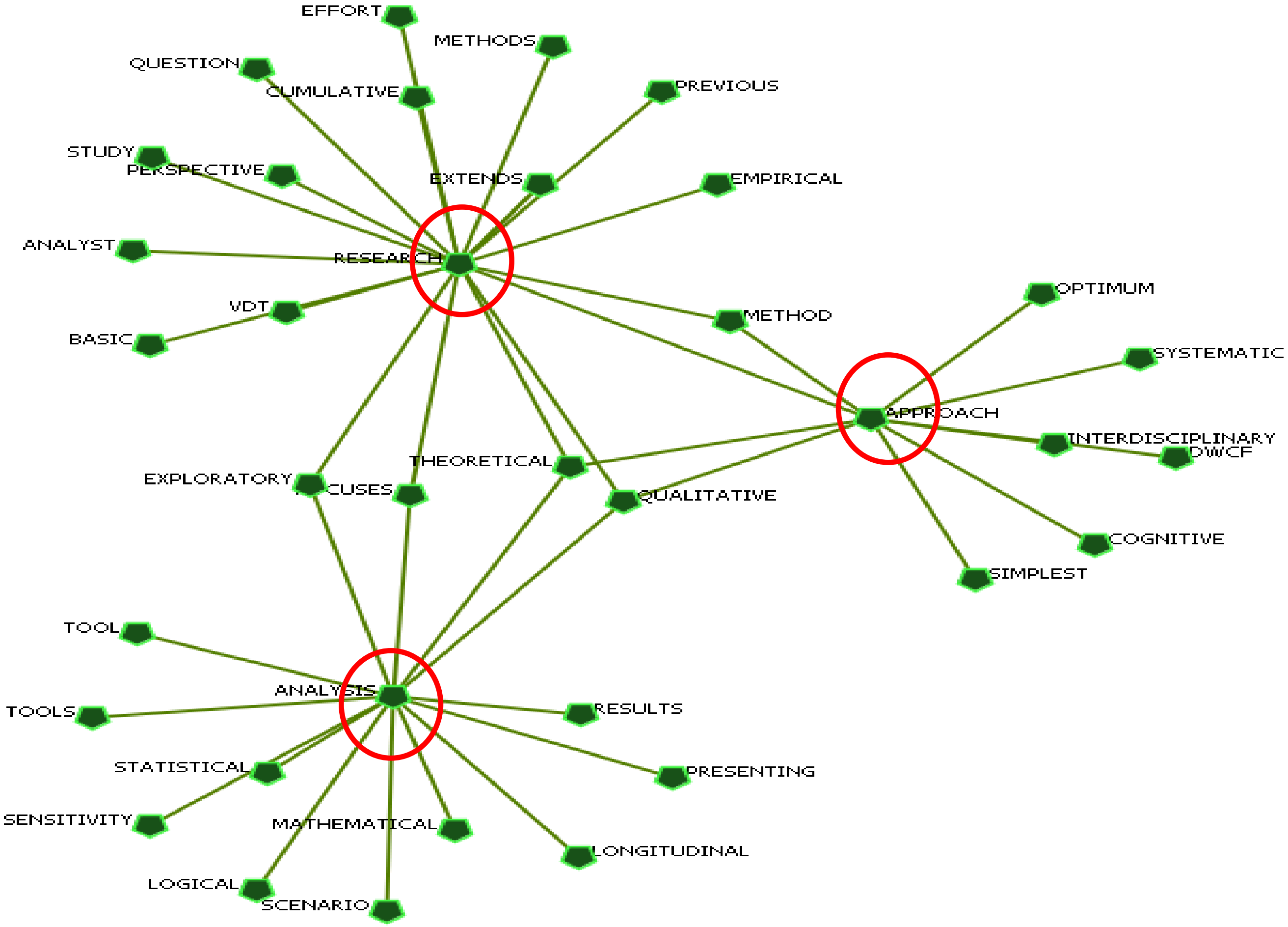
Comparing Two Systems using LLA



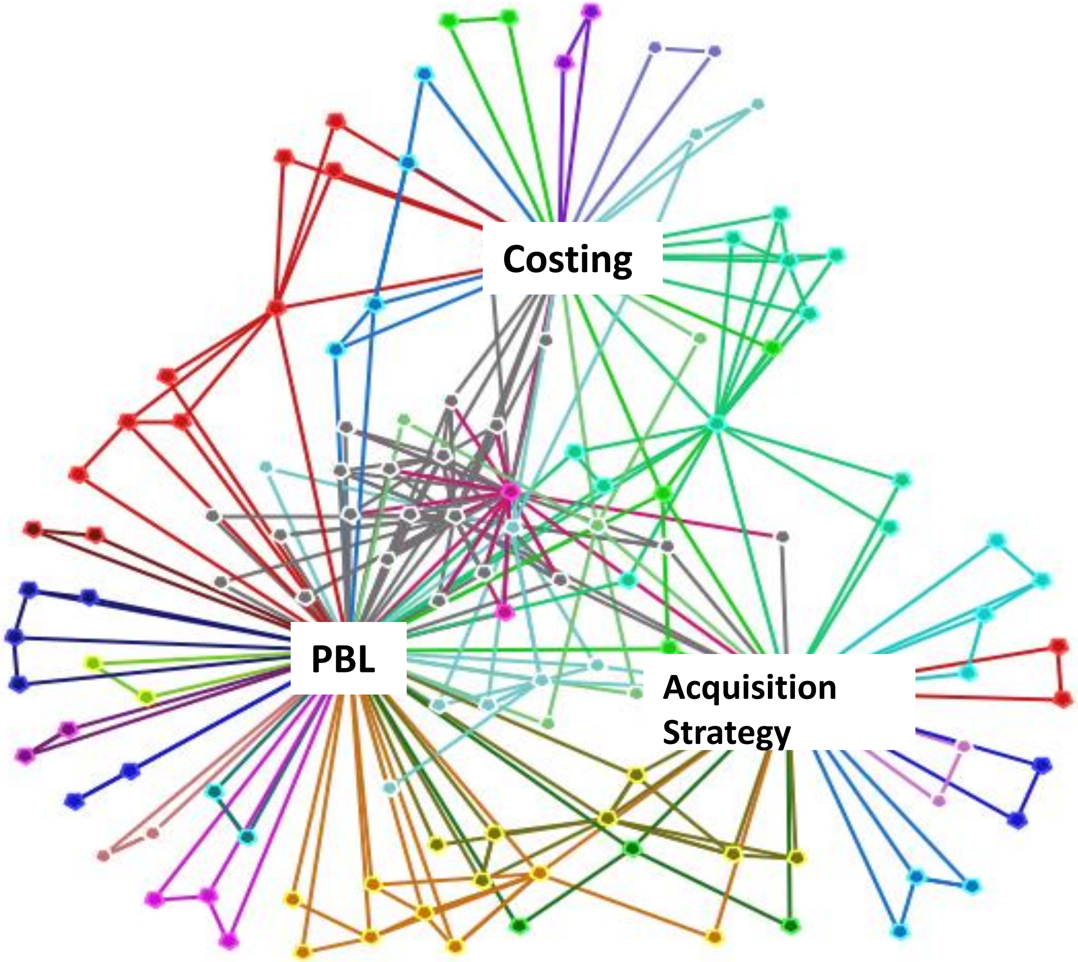
Themes



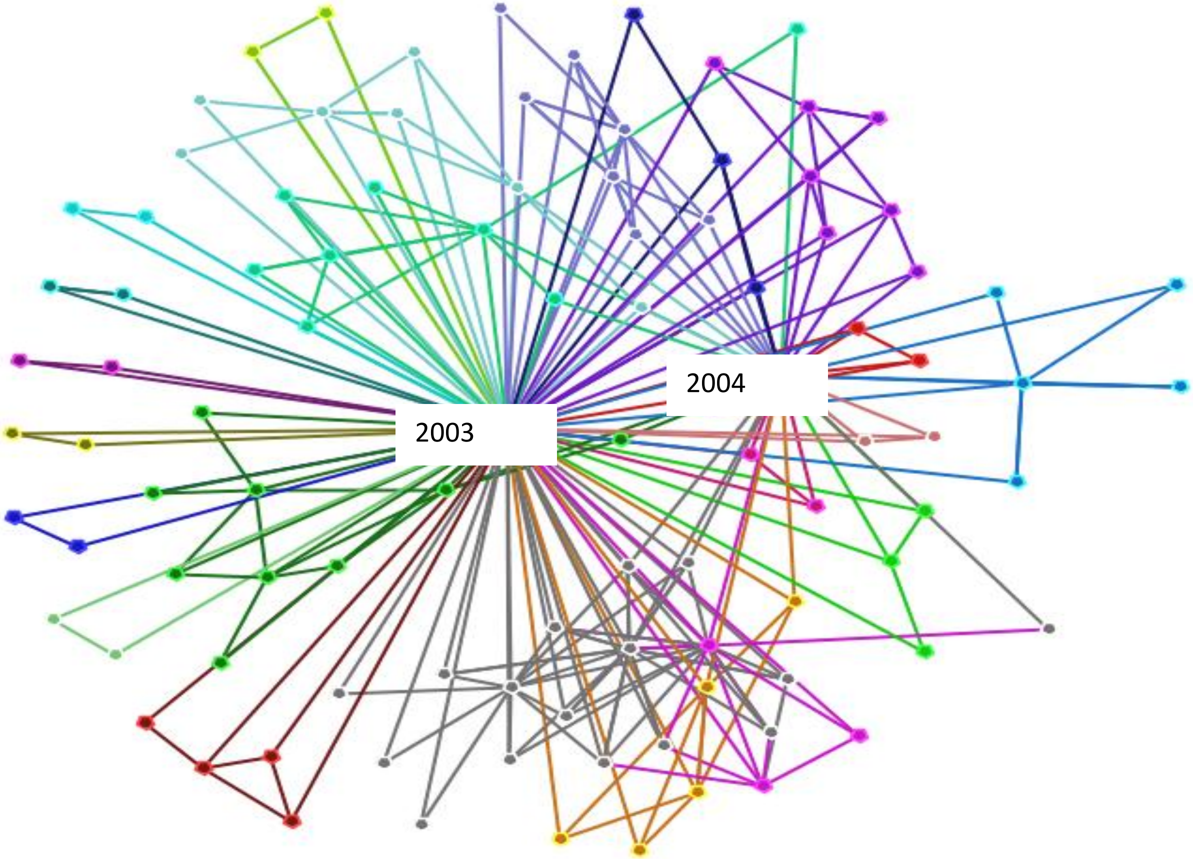
Details


EFFORT
METHODS
QUESTION
CUMULATIVE
PREVIOUS
STUDY
PERSPECTIVE
EXTENDS
EMPIRICAL
ANALYST
RESEARCH
VDT
BASIC
METHOD
OPTIMUM
SYSTEMATIC
APPROACH
INTERDISCIPLINARY
DWCF
THEORETICAL
EXPLORATORY
FOCUSES
QUALITATIVE
COGNITIVE
SIMPLEST
TOOL
ANALYSIS
TOOLS
RESULTS
STATISTICAL
PRESENTING
SENSITIVITY
MATHEMATICAL
LONGITUDINAL
LOGICAL
SCENARIO

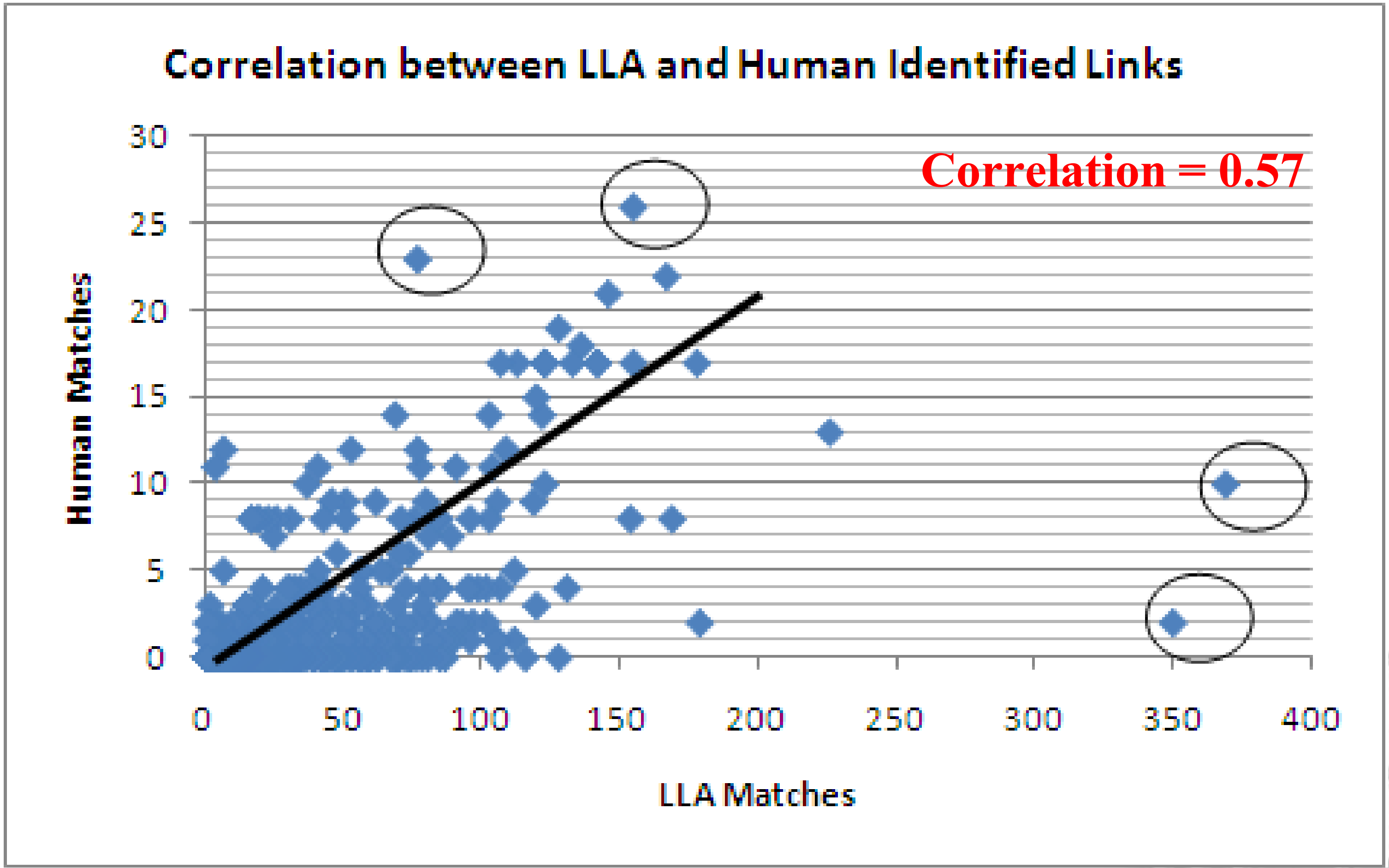
Comparing Categories



Compare Time Points



Phase I Results: Validation of LLA

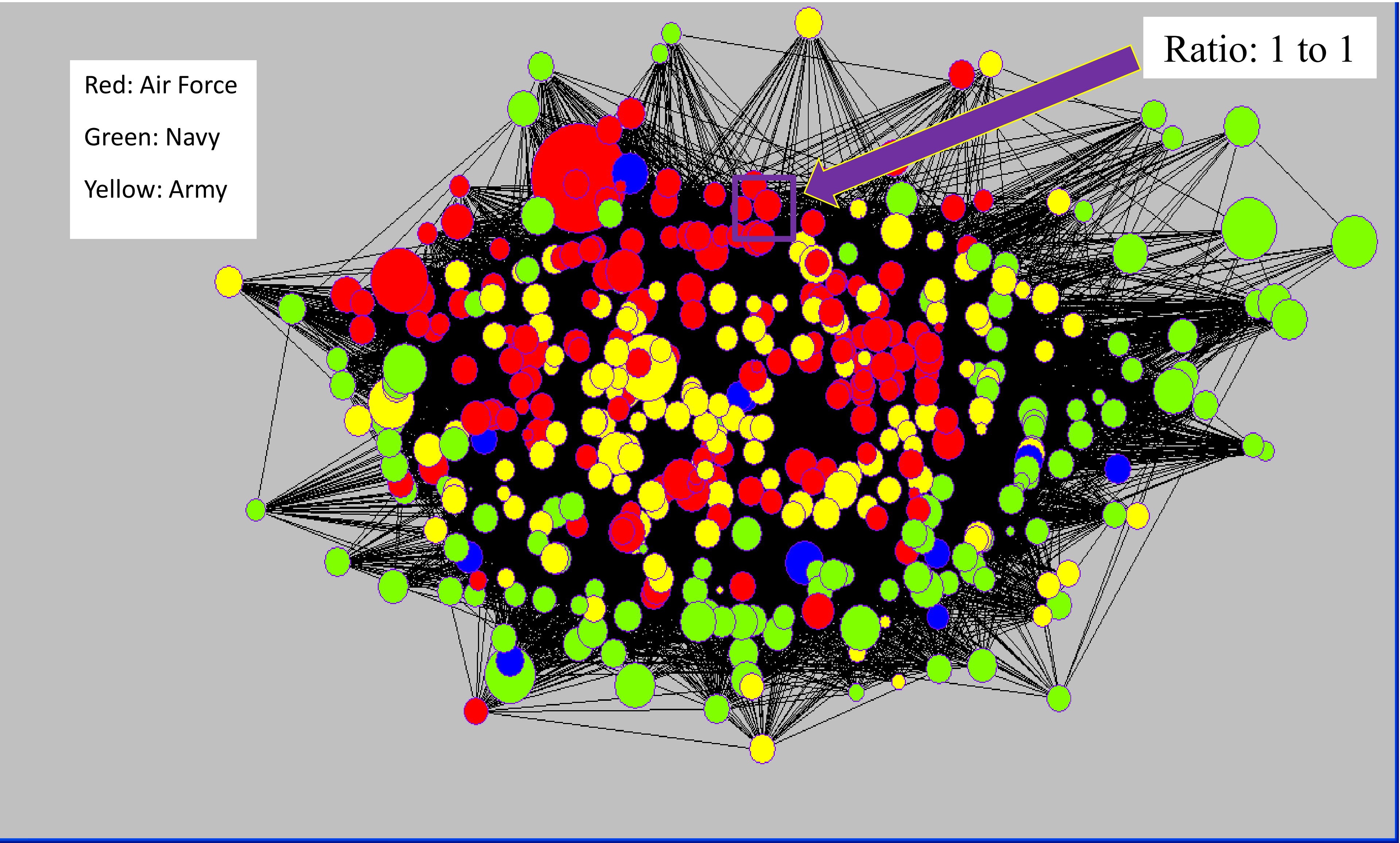


LLA Benefits

- High correlation exists between LLA results and human analyses
  - Establishes the potential to use lexical links to rank documents, concepts and themes.
- LLA can also focus on ***innovations and uniqueness*** of the analyzed documents
  - Other ranking techniques which typically sort documents based on the *popularity* or *authority, are* not based on semantics
    - E.g. PageRank by Google

Trend Analysis

Semantic Network: Size of Nodes = 2009 Cost / 2008 Cost



Phase III Objectives

- Build at least two use cases of applications of Lexical Link Analysis Web Service for large-scale automation, validation, discovery, visualization, and real-time program awareness.
- Demonstrate the methodology for assisting the DoD-wide effort of integrating and maintaining authoritative and accurate acquisition data services in both legacy and new platforms.

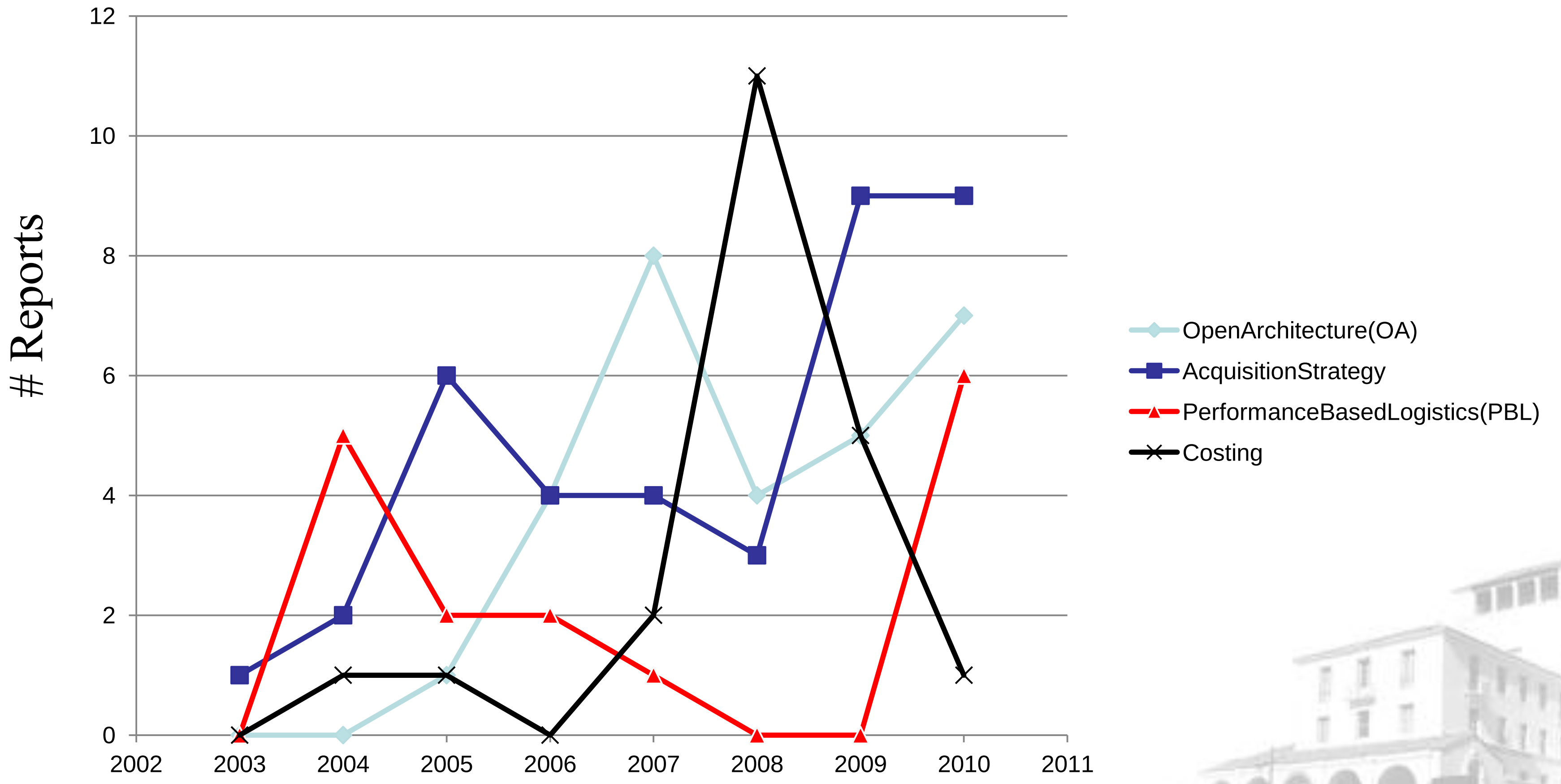
Acquisition Research Program

- 740 publications (from 2003 to 2010) from the website http://www.acquisitionresearch.net
- Pre-defined categories
  - "There are ~160 categories, e.g. Acquisition Strategy, Costing, Open architecture, Systems of Systems

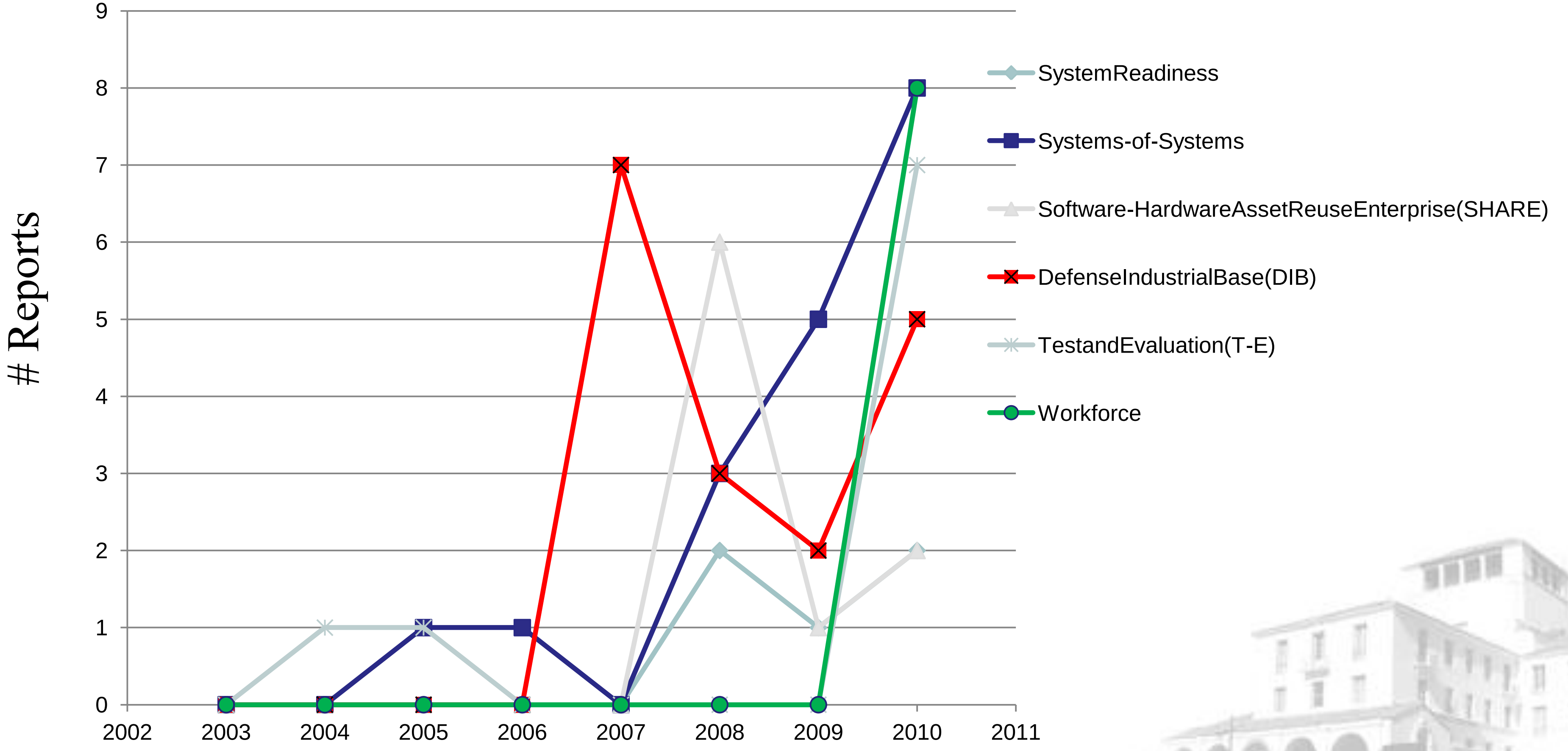
Year	# of Reports	# of Categories
2003	8	6
2004	27	17
2005	61	34
2006	62	29
2007	143	63
2008	144	68
2009	127	61
2010	184	65

ARP Reports from 2003 to 2010

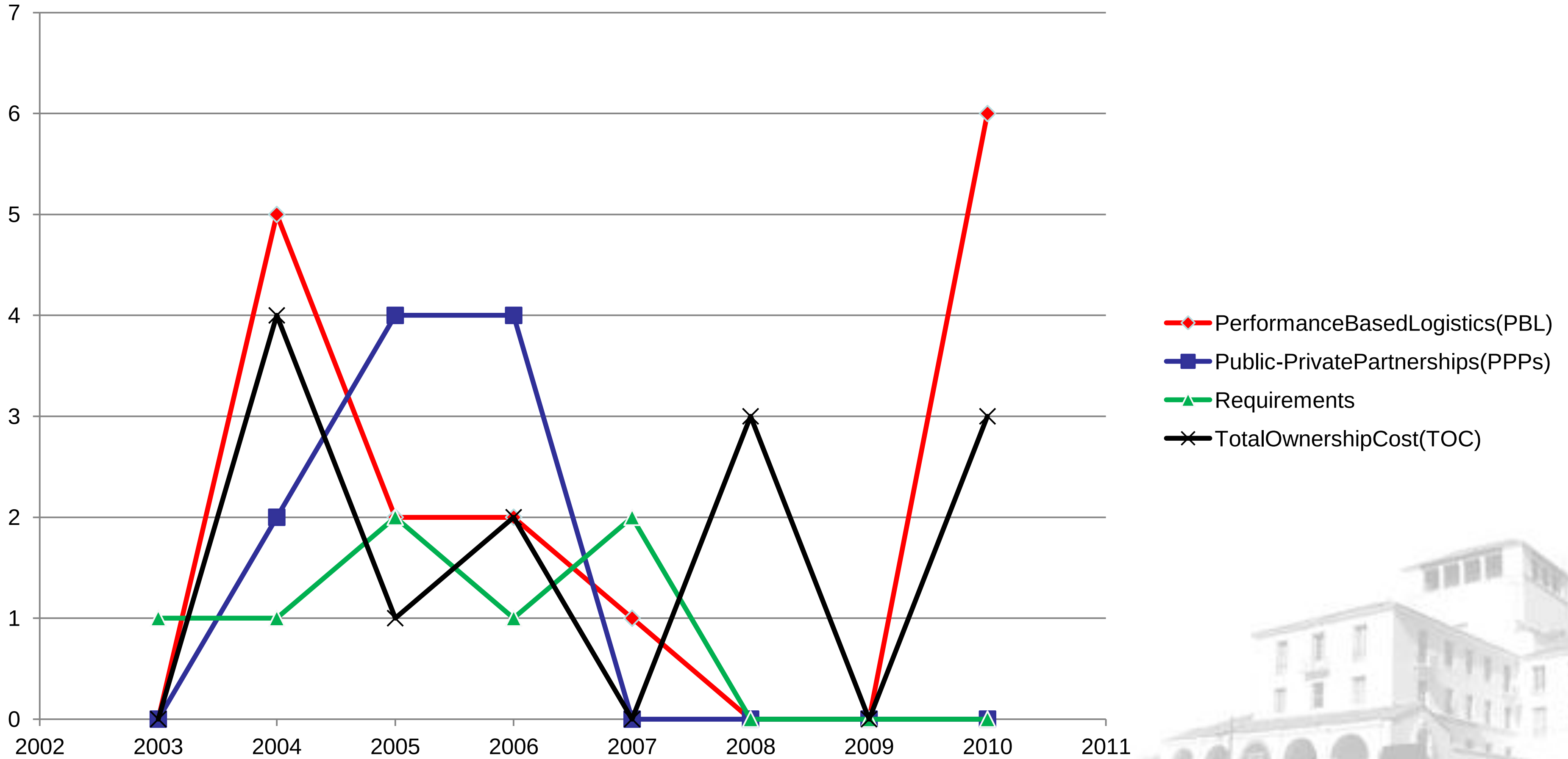
Steady Categories



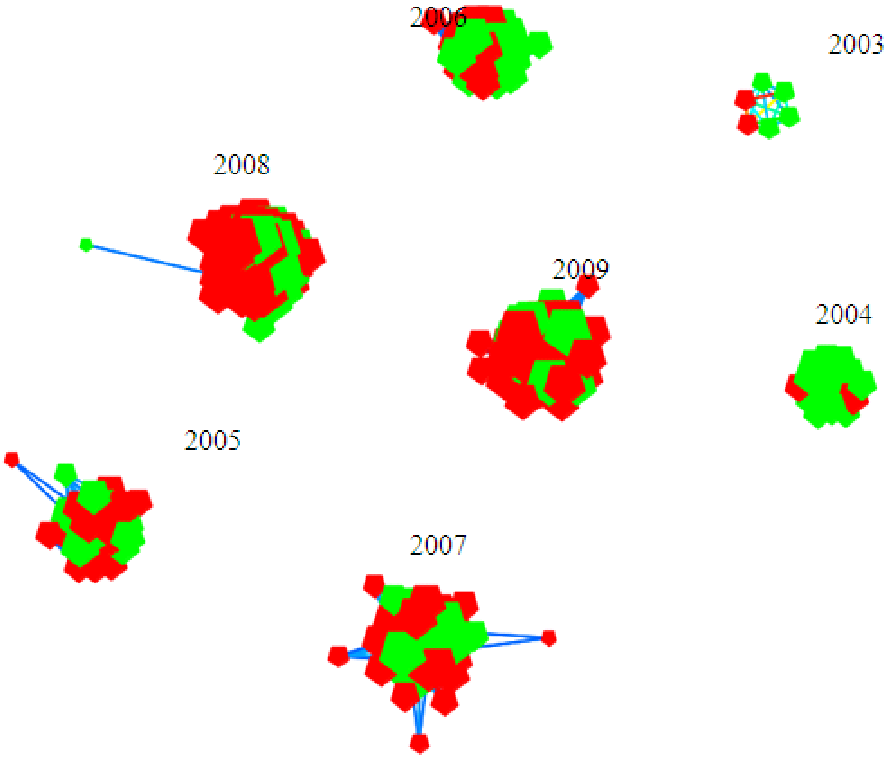
New and Emerging Categories



“Sunset” Categories

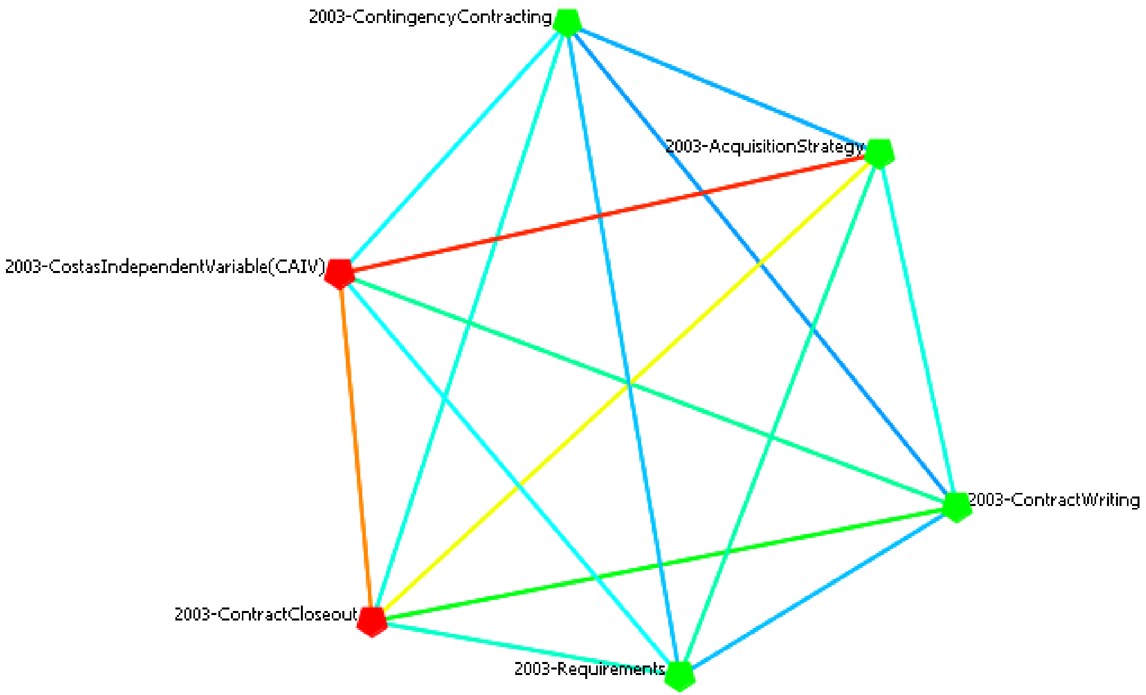


Details



- 240 objects (combinations), e.g. 2003-AcquisitionStrategy and 2004-Outsourcing,.
- For each combination
  - Label 1 (*kept*), if the associated category was continued in the following year, e.g. 2003-AcquisitionStrategy are both 2004-AcquisitionStrategy is also one
  - Label 0(*deleted*), if the associated category was not continued in the next year, e.g. 2003-ContractCloseout is an existing category, but 2004-ContractCloseout is not -- no reports were classified in the ContractCloseout category in 2004
- Semantic networks for each year
  - Green – 1(kept)
  - Red – 0 (deleted)

2003



Increased (growth, green)

- Acquisition Strategy
- Contract Writing
- Requirements
- Contingency Contracting

Reduced

Decreased (red)

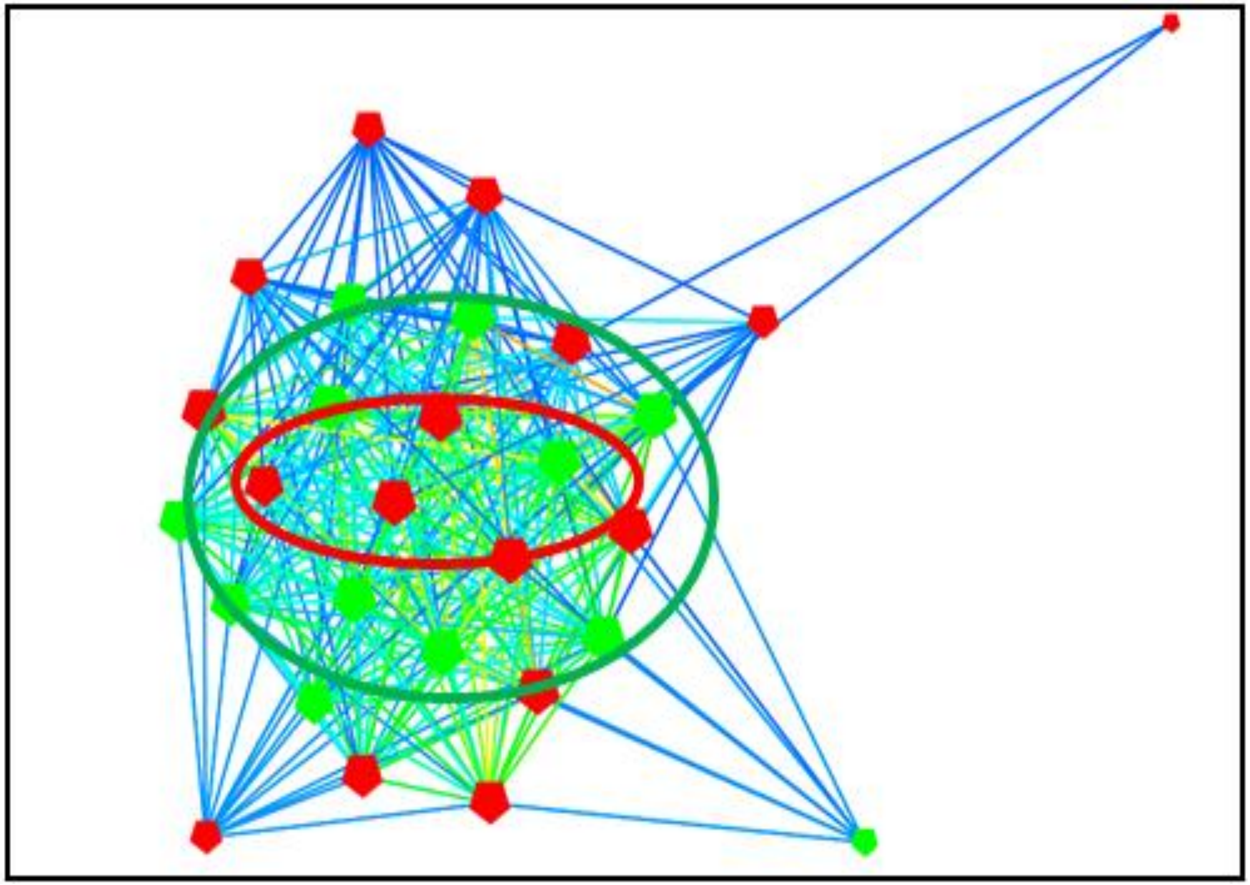
- Cost Independent Variable
- Contract Closeout

Statistical Significant Tests

	Total	Deleted	Kept	Kept/Total	
Group A (LLA Score<7)	76	53	23	0.30	
Group B (LLA Score>=7)	169	84	85	0.50	p=0.0017
Group C (Top Ranked in Total Degree)	76	47	29	0.38	
Group D Rest	169	90	79	0.47	p=0.1053

- **Green nodes have stronger (LLA scores higher) but fewer links (Total degrees lower)**

Ring of Emergence

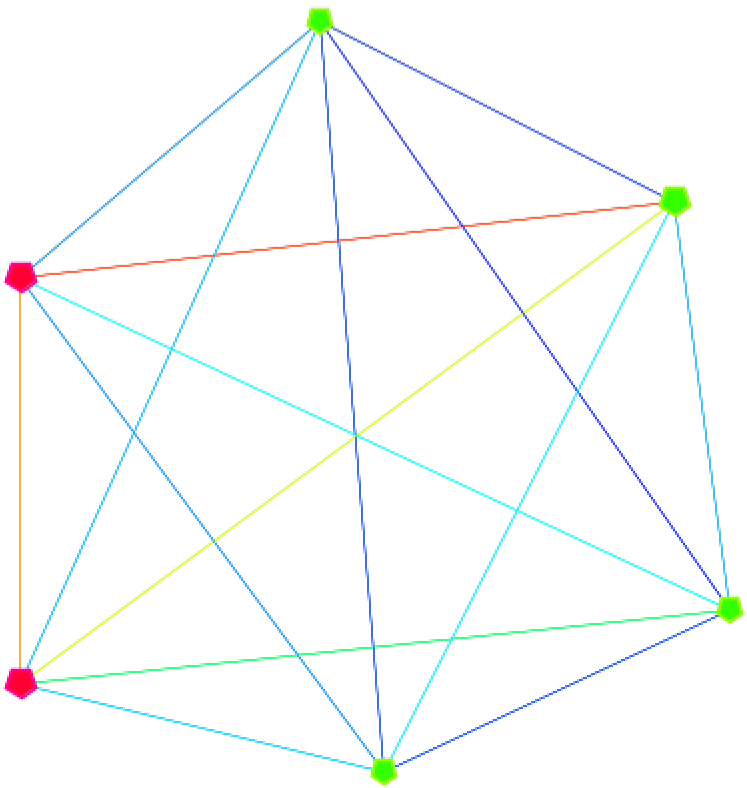


Green nodes have stronger (LLA scores higher) but fewer links (Total degrees lower)

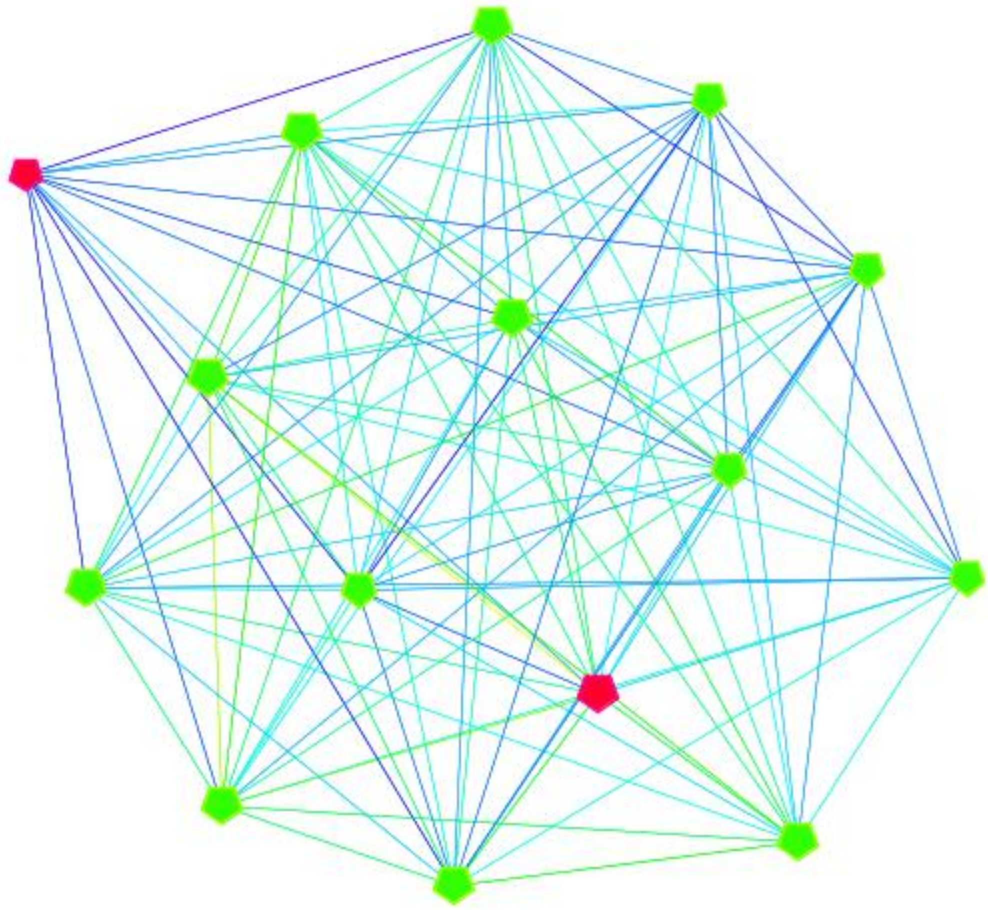
- Green nodes not in the centers but in a ring

•Associate with hotter nodes (less blue)

2003

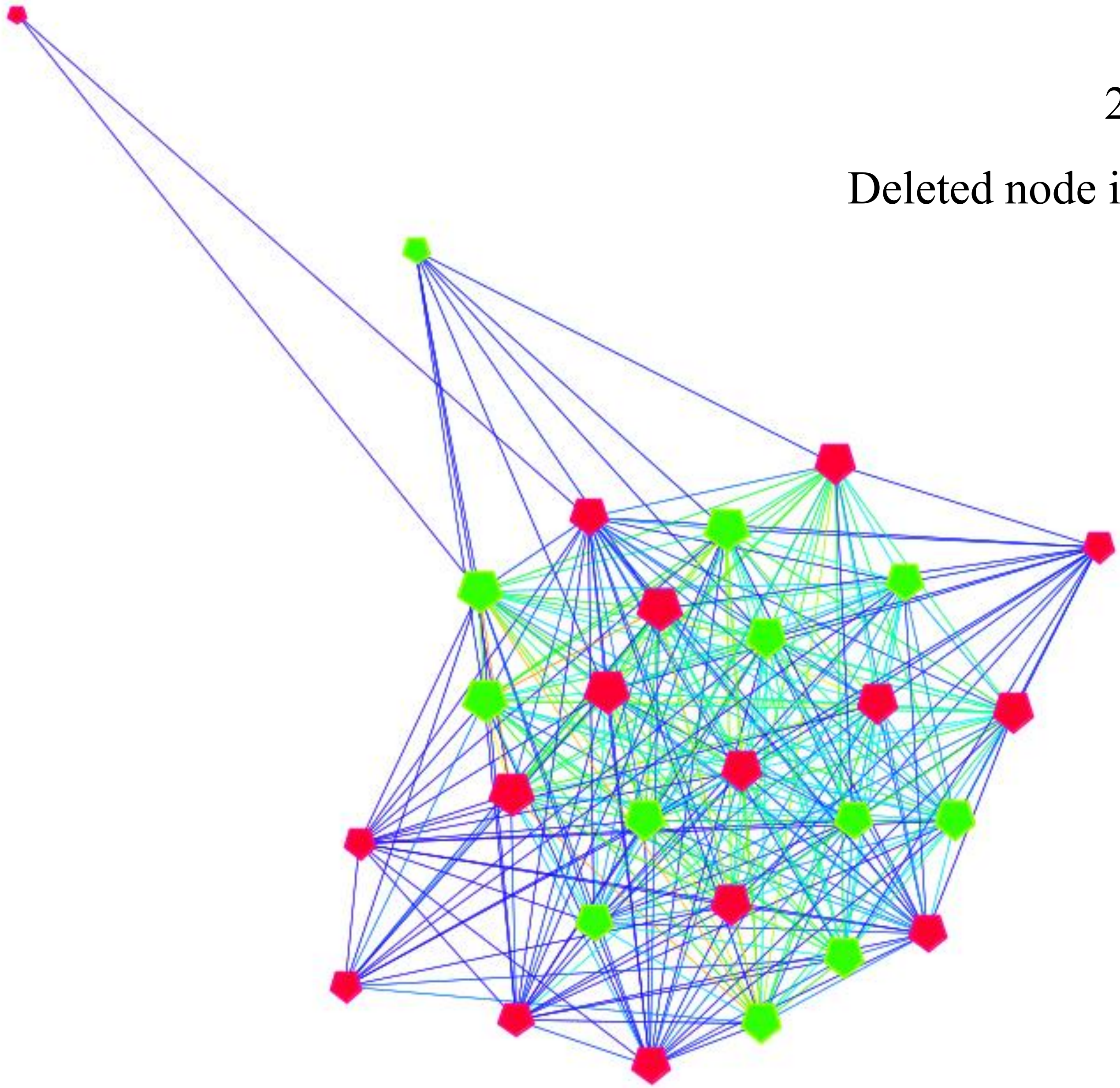


2004



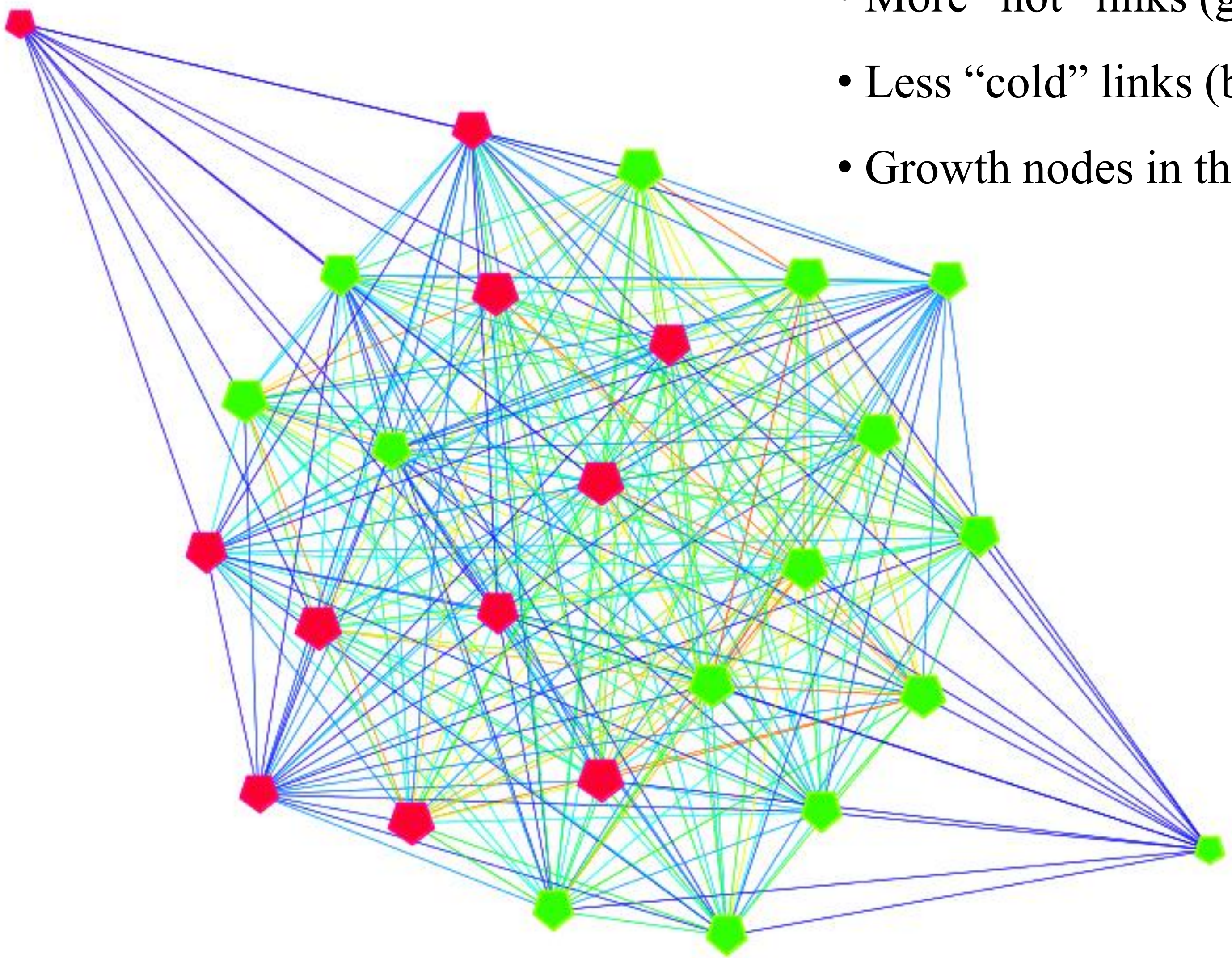
2005

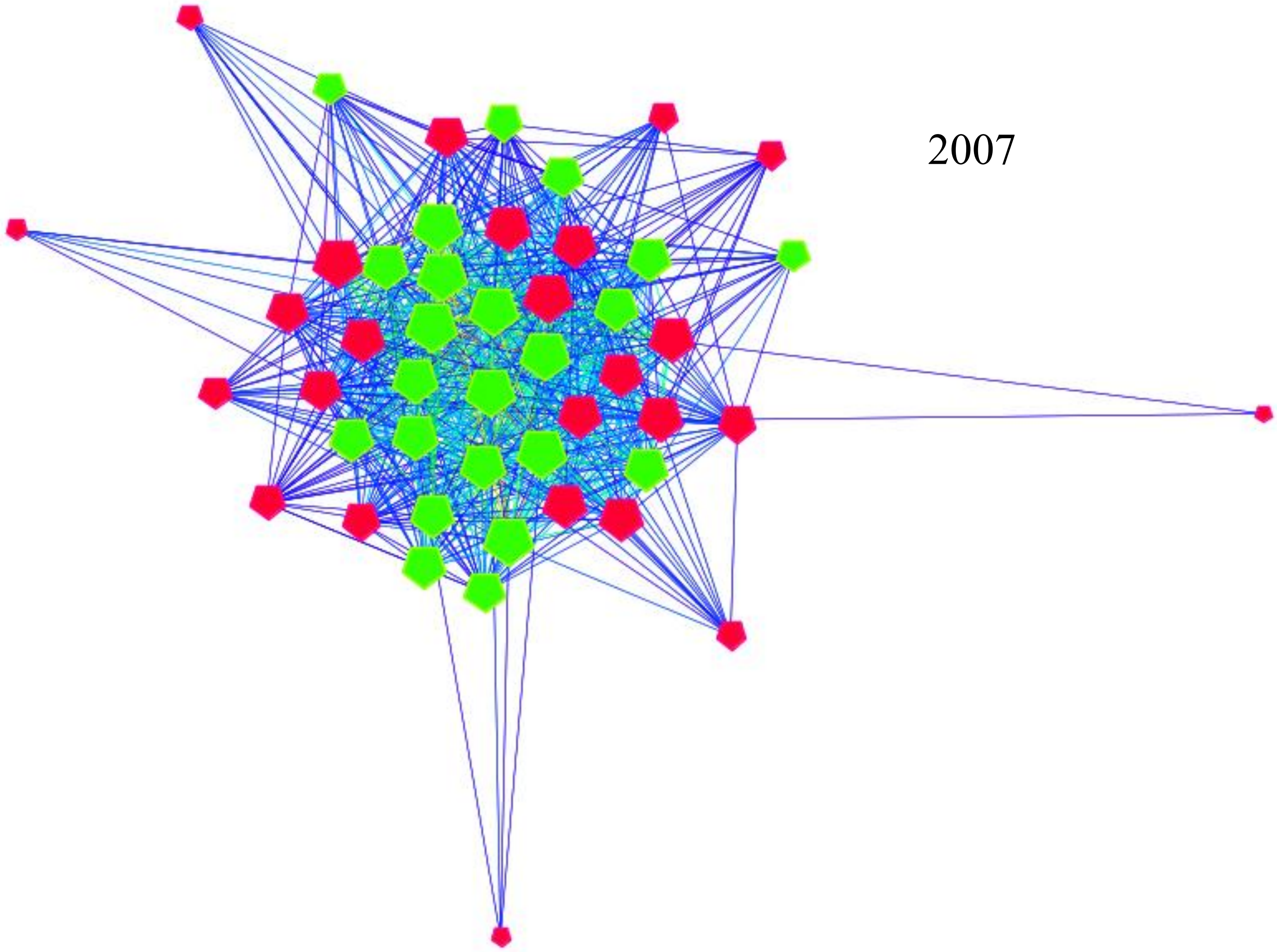
Deleted node in the "cold" areas



2006: More kept nodes (red) than deleted

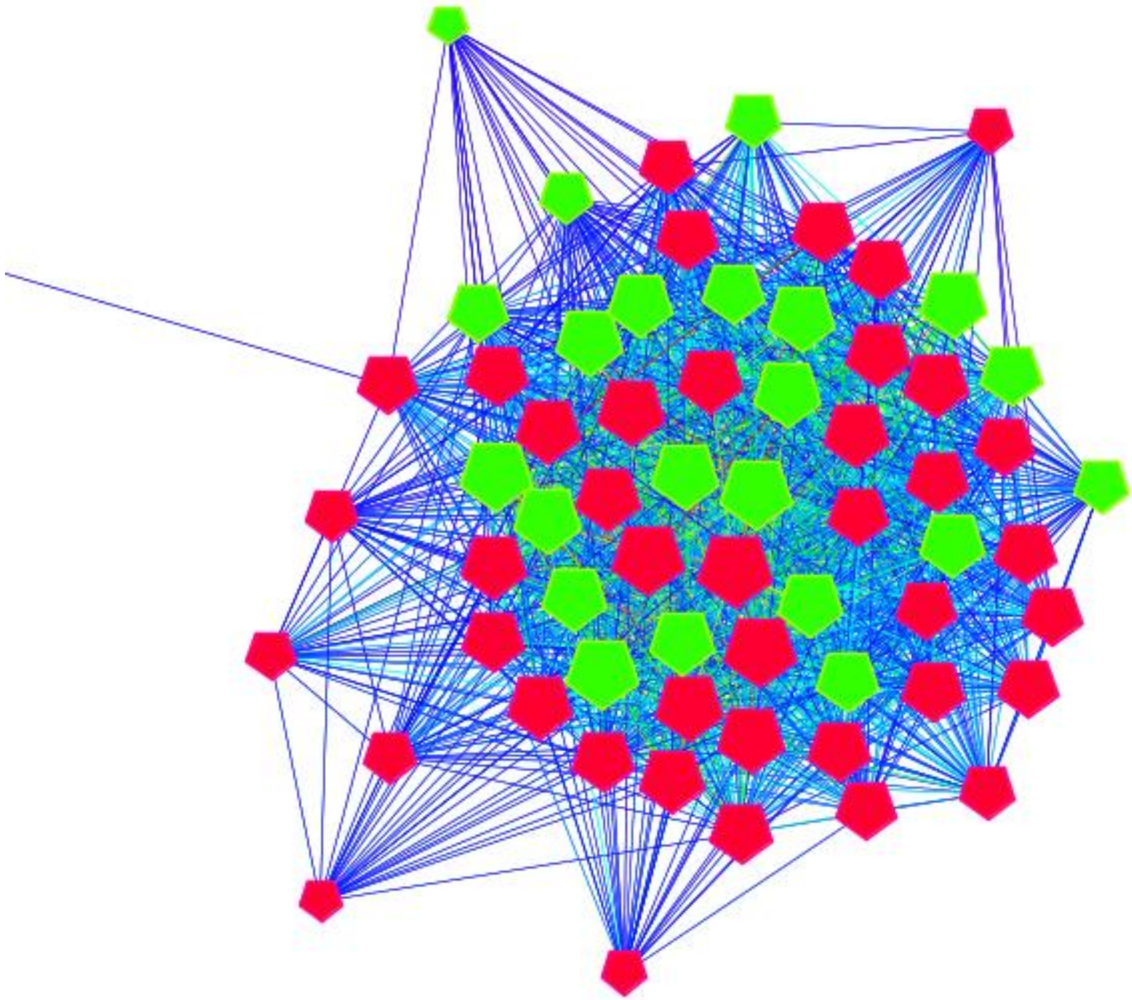
- More "hot" links (green and red)
- Less "cold" links (blue)
- Growth nodes in the "hot link" areas

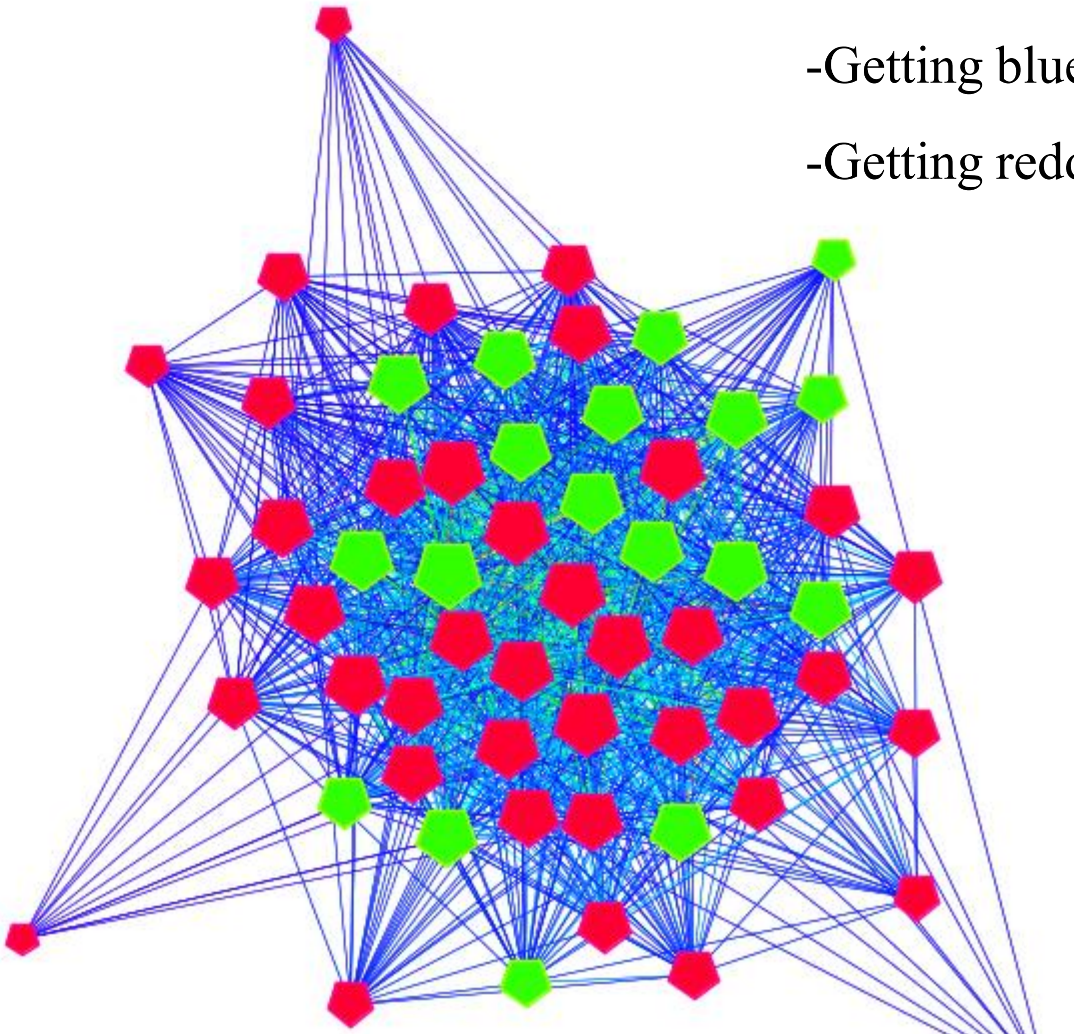




2007

2008





2009

-Getting bluer: smaller LLA scores

-Getting redder: more deleted nodes

Future Work and Why It is Important

- Is the DoD ARP system Pareto efficient?
  - How to use LLA and Collaborative Learning Agents (CLA) to make decisions that achieve an overall more efficient system
    - E.g. a DOD acquisition search system that can reinforce the diversity, uniqueness, and innovations of the technologies and investments, not just based on authorities, popularities. This could lead to a more Pareto efficient or *swarm intelligent* selection of acquisition programs

Seeking to Work with ARP Partners

- Accurate and authoritative data services in both legacy and new platforms into strategic decision-making knowledge

1. PEs: http://www.dtic.mil/descriptivesum/
2. MDAPs & ACATIIs: http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf
http://www.fas.org/man/dod-101/sys/land/wsh2007/index.html
http://www.acq.osd.mil/ara/am/sar/
3. UJTLs: http://www.dtic.mil/doctrine/jel/cjcsd/cjcsm/m350004d.pdf

- According to the Enterprise Information & OSD Studies, Office of the Under Secretary of Defense - Acquisition, Technology & Logistics (OUSD AT&L), these data sources provide the DoD-wide acquisition community with authoritative and accurate data services among others such as DAMIR(http://www.acq.osd.mil/damir/), ARA(http://www.acq.osd.mil/ara, and Selected Acquisition Report (SAR) (http://www.acq.osd.mil/ara/am/sar).

APPLICATIONS OF LEXICAL LINK ANALYSIS WEB SERVICE FOR LARGE-SCALE AUTOMATION, VALIDATION, DISCOVERY, VISUALIZATION AND REAL-TIME PROGRAM-AWARENESS

May 16-17, 2012

Dr. Ying Zhao, Dr. Douglas J. MacKinnon, Dr. Shelley P. Gallup
Research Associate Professors
Distributed Information Systems Experimentation, Naval Postgraduate School

BACK-UP SLIDES

Statistical Test Example: QAP Correlation
Quadratic Assignment Procedure [QAP; Hubert & Schultz, 1976]

QAP Correlations

		1 lla_n	2 lla_n	3 lla_n	4 lla_n	5 lla_n	6 lla_n	7 lla_n	8 lla_n
1	lla_network_1_2010-AcquisitionStrategy	1.000	0.174	0.156	0.155	0.036	0.111	0.020	0.062
2	lla_network_1_2003-AcquisitionStrategy	0.174	1.000	0.447	0.149	0.052	0.119	0.043	0.089
3	lla_network_1_2004-AcquisitionStrategy	0.156	0.447	1.000	0.111	0.047	0.119	0.051	0.080
4	lla_network_1_2005-AcquisitionStrategy	0.155	0.149	0.111	1.000	0.156	0.084	0.034	0.088
5	lla_network_1_2006-AcquisitionStrategy	0.036	0.052	0.047	0.156	1.000	0.067	0.036	0.056
6	lla_network_1_2007-AcquisitionStrategy	0.111	0.119	0.119	0.084	0.067	1.000	0.097	0.123
7	lla_network_1_2008-AcquisitionStrategy	0.020	0.043	0.051	0.034	0.036	0.097	1.000	0.286
8	lla_network_1_2009-AcquisitionStrategy	0.062	0.089	0.080	0.088	0.056	0.123	0.286	1.000

QAP P-Values

		1 lla_n	2 lla_n	3 lla_n	4 lla_n	5 lla_n	6 lla_n	7 lla_n	8 lla_n
1	lla_network_1_2010-AcquisitionStrategy	0.000	0.020	0.020	0.020	0.020	0.020	0.020	0.020
2	lla_network_1_2003-AcquisitionStrategy	0.020	0.000	0.020	0.020	0.020	0.020	0.020	0.020
3	lla_network_1_2004-AcquisitionStrategy	0.020	0.020	0.000	0.020	0.020	0.020	0.020	0.020
4	lla_network_1_2005-AcquisitionStrategy	0.020	0.020	0.020	0.000	0.020	0.020	0.020	0.020
5	lla_network_1_2006-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.000	0.020	0.020	0.020
6	lla_network_1_2007-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.000	0.020	0.020
7	lla_network_1_2008-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.020	0.000	0.020
8	lla_network_1_2009-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.000

QAP statistics saved as datafile QAP Correlation Results

Program Elements: Center of Many Things

http://www.dtic.mil/descriptivesum/

UNCLASSIF[illegible]

Exhibit R-2a, RDT&E Project Justifica[illegible]

DATE: May 2009

BUDGET ACTIVITY	PE NUMBER AND TITLE	PROJECT NUMBER AND TITLE
05 System Development and Demonstration (SDD)	0604602F Armament/Ordnance Development	5361 Stores-Aircraft Interface

Cost ($ in Millions)	FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total
5361 Stores-Aircraft Interface	0.000	0.000	6.685	[illegible]	0.000	0.000	0.000	0.000	Continuing	TBD
Quantity of RDT&E Articles	0	0	0	0	0	0	0	0		

In FY 2010, Project 5361, Stores-Aircraft Interface (new), efforts were transferred from PE 0605011F, RDT&E for Aging Aircraft, Project 654685, Universal Armament Interface (UAI), in order to properly fund the maturing technology.

(U) **A. Mission Description and Budget Item Justification**

Universal Armament Interface (UAI) is an Air Force initiative to develop, enhance, and implement standardized interfaces in aircraft, weapons and mission planning to support integration of weapons independent of aircraft Operation Flight Program (OFP) cycles. UAI is currently being implemented on the F-15E and F-16 Block 40/50 aircraft, Small Diameter Bomb (SDB) I and II, Joint Direct Attack Munition (JDAM), Joint Air-to-Surface Stand-off Missile (JASSM) and Precision Guided Munitions Planning Software (PGMPS). Additional aircraft and weapons have program plans to implement UAI. The UAI program office is responsible for development and enhancement of the standard, provision of certification tools (test assets) and implementation support to aircraft and weapons.

The UAI efforts were transferred (1) to ensure continued funding for UAI through the FYDP (PE 0605011F will be zeroed out in FY 2010 due to higher Air Force priorities), and (2) to properly fund the maturing technology. The new project number is established to provide greater visibility into UAI's budget. Funding UAI via the Arm/Ord PE will ensure that platform and weapon program offices have the support required to implement and update UAI.

This program is in Budget Activity 5 - System Development and Demonstration (SDD) because it supports armament integration, an SDD-type activity.

0604602F references 0605011F Forward Link

0605011F referenced by 0604602F Backward Link

	(U) B. Accomplishments/Planned Program ($ in Millions)	FY 2008	FY 2009	FY 2010
(U)	ICD Dev/Updates			5.702
(U)	UAI Common Component			[illegible]
(U)	Certification Tool			0.197
(U)	Total Cost	0.000	0.000	6.685

This is not a new start; these efforts were performed under PE 0605011F, RDT&E for Aging Aircraft, in FY 2008 and FY 2009.

(U) **C. Other Program Funding Summary ($ in Millions)**

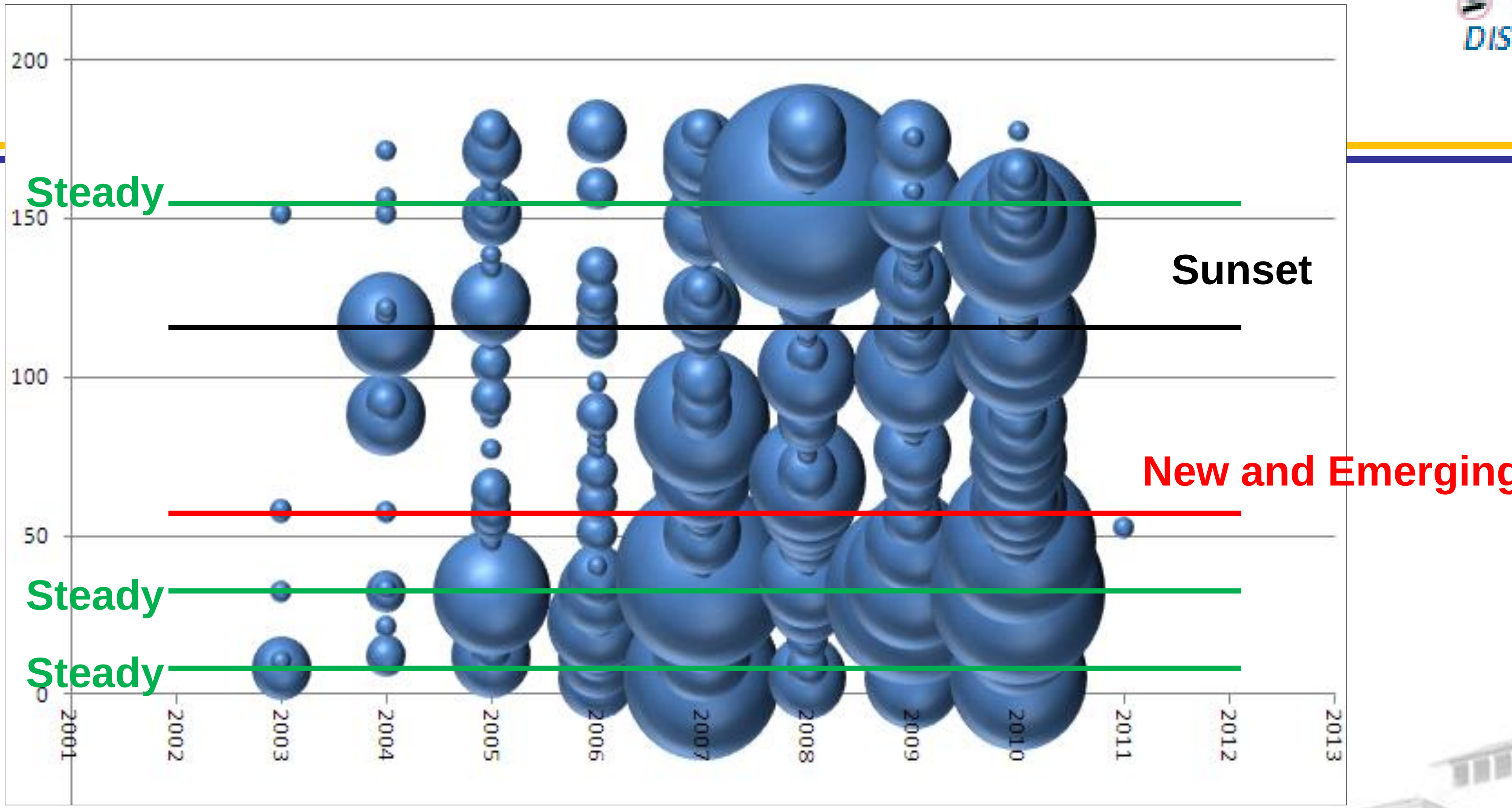
	FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total Cost
(U) N/A										

(U) **D. Acquisition Strategy**

In December 2004, under the authority of a class Justification and Approval (J&A), the UAI program office awarded individual Cost Plus Fixed Fee (CPFF) contracts to Boeing, Lockheed-Martin, Northrop-Grumman and Raytheon. These four vendors are the Original Equipment Manufacturers (OEMs) for approximately 90% of the Department of Defense' platforms and weapons. Each OEM is responsible for a different piece of the total UAI requirement based on its platform or weapon expertise.

Statistical Significance Tests (Pre-defined Categories)

	Centrality Authority	Radials	Simmelian Ties	Centrality Total Degree	Triad Count	Rank	Value
Growth	0.732	0.481	0.123	0.415	1967.766	2.481	1.104
Die-out	0.665	0.278	0.150	0.478	2646.340	1.423	-1.799
p-value	0.015	0.0015	<0.0001	0.028	0.0002		



•Steady categories in which the number of reports increased

•New and emerging categories in which there were relatively new.

•Die-down categories in which the number of reports reduced.

Apply LLA to Understand Why Categories Steady, Emerging and Disappearing

- Object: a Year-Category combination
- Link: LLA Score of overlaps of reports for the year and category

Automatic Categories

- Apply LLA to automatically generate themes combined with years as categories
  - 225 of such automatic categories
    - E.g. 2003-COST*COSTS*TOTAL & 2004-SYSTEMS*SYSTEM*PROGRAM
  - We define a value of an automatic category as
    - # of lexical links in the time frame for the theme –

    # of lexical links in the time frame for the same theme
  - Compute the centrality measures for the 225 nodes
    - Links only computed within the same time frame
  - Compute correlation between the centrality measures and "values" of the nodes

e.g. Correlation between "Centrality Authority" and "Value" =0.23 (p<0.05 n=225)

Node ID	Centrality Authority/knowledge x know	Rank	Value
2004-SYSTEMS*SYSTEM*PROGRAM	0.9758	3	94
2004-PERSONNEL*MILITARY*SUPPORT	1	3	90
2004-BUSINESS*INDUSTRY*ARMY	0.7449	3	14
2004-COST*COSTS*TOTAL	0.2685	3	74
2004-CONTRACT*PERFORMANCE*CONTRACTS	0.622	3	29
2003-MODEL*ANALYSIS*APPROACH	0.8503	3	22
2003-PERSONNEL*MILITARY*SUPPORT	0.7443	3	22
2003-SYSTEMS*SYSTEM*PROGRAM	0.525	3	9
2003-PROCESS*PROCESSES*PHASE*PLANNING	1	2	1
2004-SOFTWARE*COMPONENTS*ENGINE*POWER	0.5268	3	25
2004-MANAGEMENT*DECISION*REVIEW	0.3725	3	16

Automatically generated categories

Statistical Significant Correlations Between Centrality and Growth

Pearson Correlation	Centrality Authority (Eigenvalue,PageRank)	Centrality Betweenness	Correlation Expertise	Correlation Resemblance	Centrality Total Degree	Triad Count	Samples	p-value
ARP automatic	0.23	0.24	0.19				225	<0.05
ARP categories			0.15	0.18	-0.12	-0.17	272	<0.05

*Empty cells mean the correlations are not statistically significant

Sort by "Centrality Authority"

arp_automatic_2_cost.xlsx - Microsoft Excel

	A	C	AZ
1	Node ID	Centrality Authority/knowledge x knowledge	Value
2	2007-SYSTEMS*SYSTEM*PROGRAM	1	153
3	2008-SYSTEMS*SYSTEM*PROGRAM	1	-86
4	2009-SYSTEMS*SYSTEM*PROGRAM	1	35
5	2004-PERSONNEL*MILITARY*SUPPORT	1	90
6	2005-SYSTEMS*SYSTEM*PROGRAM	1	25
7	2003-PROCESS*PROCESSES*PHASE*PLANNING	1	1
8	2006-SYSTEMS*SYSTEM*PROGRAM	1	18
9	2004-SYSTEMS*SYSTEM*PROGRAM	0.9758	94
10	2008-PERSONNEL*MILITARY*SUPPORT	0.9258	-27
11	2009-PERSONNEL*MILITARY*SUPPORT	0.9011	48
12	2007-COST*COSTS*TOTAL	0.8795	17
13	2007-PERSONNEL*MILITARY*SUPPORT	0.8629	7
14	2003-MODEL*ANALYSIS*APPROACH	0.8503	22
15	2007-MODEL*ANALYSIS*APPROACH	0.8453	32
16	2003-CONTRACT*PERFORMANCE*CONTRACTS	0.8405	2
17	2009-PROCESS*PROCESSES*PHASE*PLANNING	0.8174	23
18	2007-MANAGEMENT*DECISION*REVIEW	0.8164	1
19	2009-MODEL*ANALYSIS*APPROACH	0.8076	-4
20	2006-PERSONNEL*MILITARY*SUPPORT	0.7956	39
21	2006-COST*COSTS*TOTAL	0.7649	45
22	2009-COST*COSTS*TOTAL	0.7604	31
23	2008-MODEL*ANALYSIS*APPROACH	0.7456	-6
24	2004-BUSINESS*INDUSTRY*ARMY	0.7449	14
25	2003-PERSONNEL*MILITARY*SUPPORT	0.7443	22
26	2006-BASED*PRICE*JOINT	0.7306	9
27	2003-COST*COSTS*TOTAL	0.7256	2
28	2005-PERSONNEL*MILITARY*SUPPORT	0.7173	-15
29	2005-COST*COSTS*TOTAL	0.7126	1
30	2006-MODEL*ANALYSIS*APPROACH	0.7048	69

Sort by “Correlation Expertise”

arp_automatic_2_cost.xlsx - Microsoft Excel

	A	T	AZ
1	Node ID	Correlation Expertise/knowledge x k	Value
2	2004-SYSTEMS*SYSTEM*PROGRAM	0.0329	94
3	2004-BUSINESS*INDUSTRY*ARMY	0.0328	14
4	2004-PERSONNEL*MILITARY*SUPPORT	0.0328	90
5	2004-COST*COSTS*TOTAL	0.0327	74
6	2004-CONTRACT*PERFORMANCE*CONTRACTS	0.0327	29
7	2003-SYSTEMS*SYSTEM*PROGRAM	0.0327	9
8	2003-PERSONNEL*MILITARY*SUPPORT	0.0327	22
9	2003-MODEL*ANALYSIS*APPROACH	0.0327	22
10	2003-PROCESS*PROCESSES*PHASE*PLANNING	0.0327	1
11	2004-SERVED*LCDR	0.0326	11
12	2004-MANAGEMENT*DECISION*REVIEW	0.0326	16
13	2004-SOFTWARE*COMPONENTS*ENGINE*POWER	0.0326	25
14	2003-COST*COSTS*TOTAL	0.0326	2
15	2003-REPORT*REPORTS*ACT	0.0325	1
16	2007-TRAINING*COURSES*INSTRUCTION	0.0325	17
17	2007-SERVED*LCDR	0.0325	15
18	2004-REPORT*REPORTS*ACT	0.0325	15
19	2004-AIR_FORCE*NAVY*AIR	0.0325	12
20	2004-DUE*CONTROL*INVENTORY	0.0325	21
21	2004-BASED*PRICE*JOINT	0.0325	8
22	2003-DATA*INFORMATION*KEY	0.0325	4
23	2004-TIME*SIGNIFICANT*ADDITIONAL*FUNDING	0.0325	53
24	2003-MANAGEMENT*DECISION*REVIEW	0.0325	9
25	2004-MODEL*ANALYSIS*APPROACH	0.0325	55
26	2004-ACQUISITION*DEFENSE*NATIONAL	0.0325	21
27	2003-ACQUISITION*DEFENSE*NATIONAL	0.0325	5
28	2003-PRIOR*COMPETITION*SPECIFIC	0.0325	-4
29	2003-CONTRACT*PERFORMANCE*CONTRACTS	0.0325	2

School

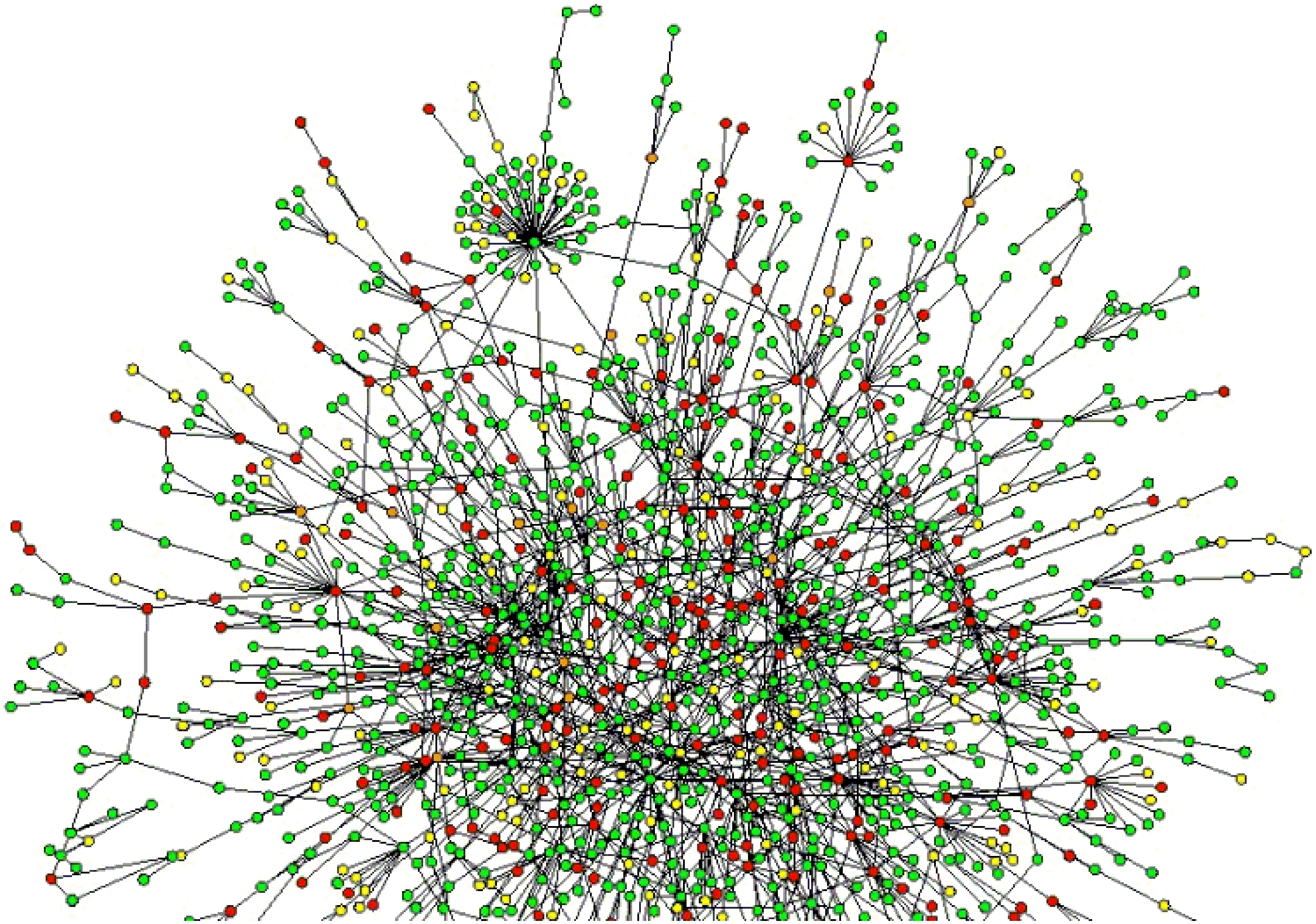
THEORY

The effect of linguistic constraints on the large scale organization of language

Madhav Krishna[1], Ahmed Hassan[2], Yang Liu[2], Dragomir Radev[2*]
1 Columbia University New York, New York, USA
2 University of Michigan Ann Arbor, Michigan, USA
* E-mail: Corresponding radev@umich.edu

- Characteristics of a set of important networks and systems of systems
  - WWW , collaboration networks, social networks, US power grid, metabolic networks, semantic networks,
  - Share the same characteristics
    - Power-law, scale-free: relatively small number of well-connected nodes serve as hubs Pareto principle, 80/20 rule
    - Small-world phenomenon (random two nodes ,e.g. two person in US, only separated by six degrees away)
    - Self-organizing
    - Self similar (fractals)
    - Preferential attachment

The E.coli metabolic network is scalefree (PZM Pareto-Zipf-Mandelbrot type, parabolic fractal) and has small-world properties



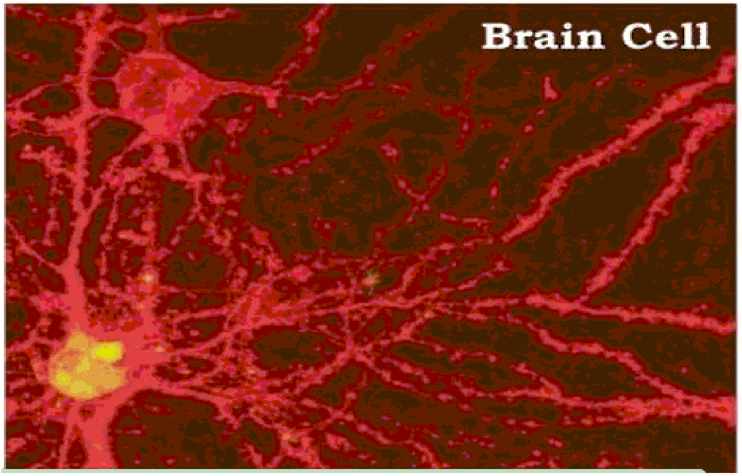
http://www.bordalierinstitute.com/targetd.html Connect to fractals?

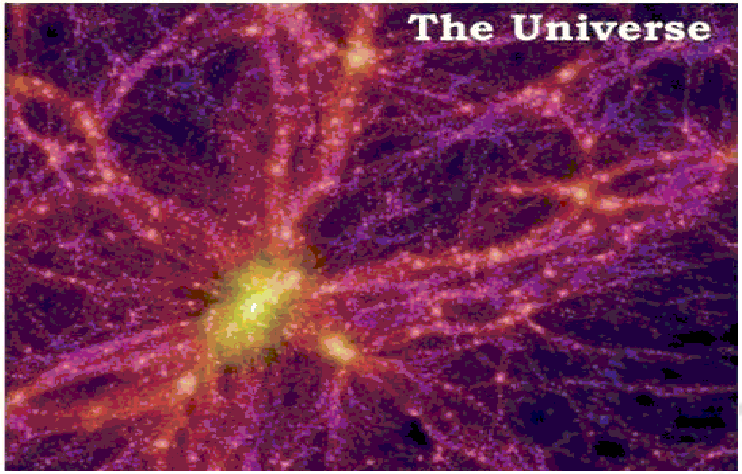
bordalier institute

contact: winiwarter@bordalierinstitute.com

Research scope : complex systems / neural networks & evolution

"I think the next century (21st) will be the century of complexity." Stephen Hawking





en.wikipedia.org/wiki/File:Water_Crystals_on_Mercury_20Feb2010_CU1.jpg

DIA
lopedia

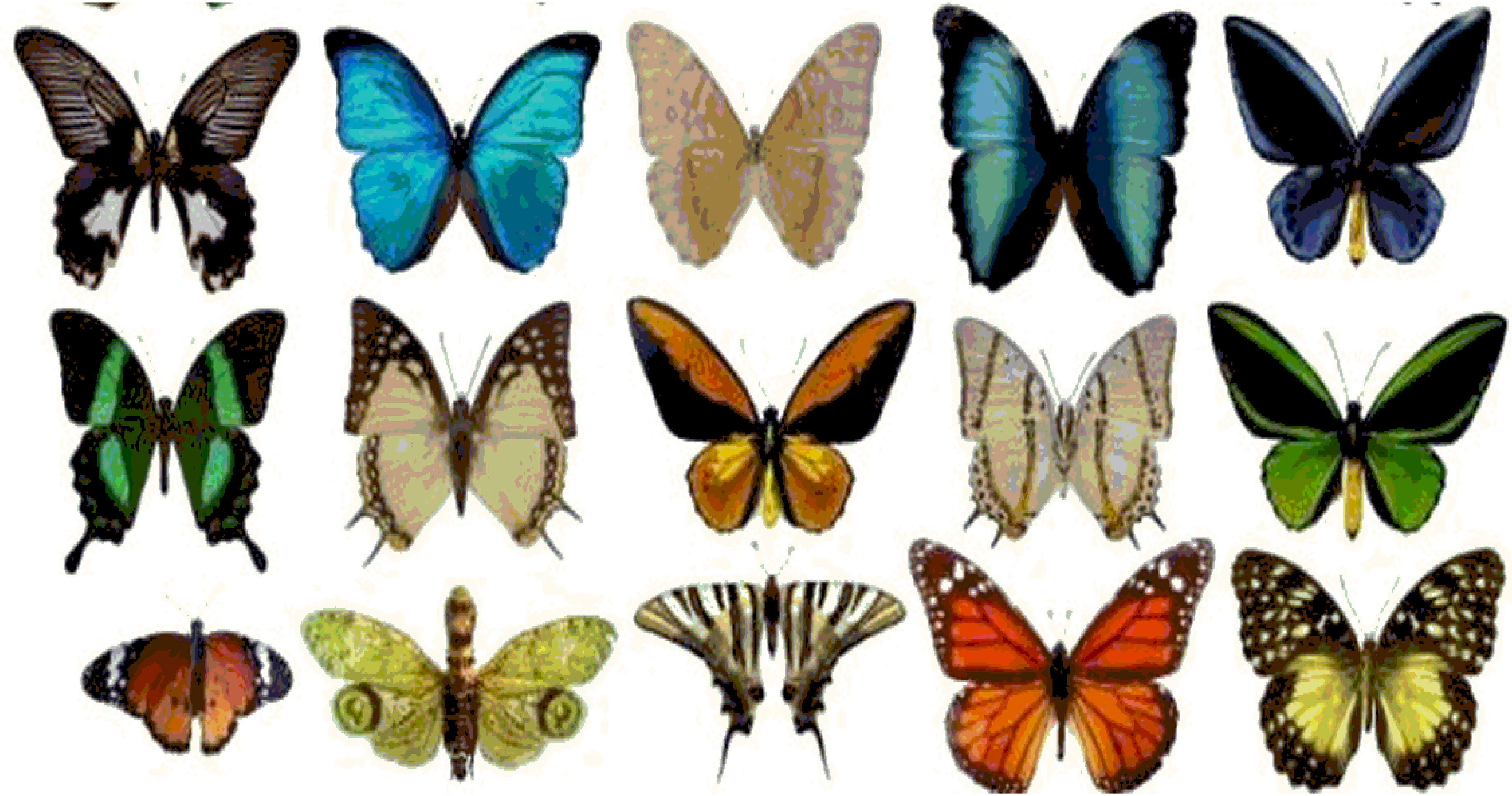
File Talk

File:Water Crystals on Mercury 20Feb2010 CU1.jpg

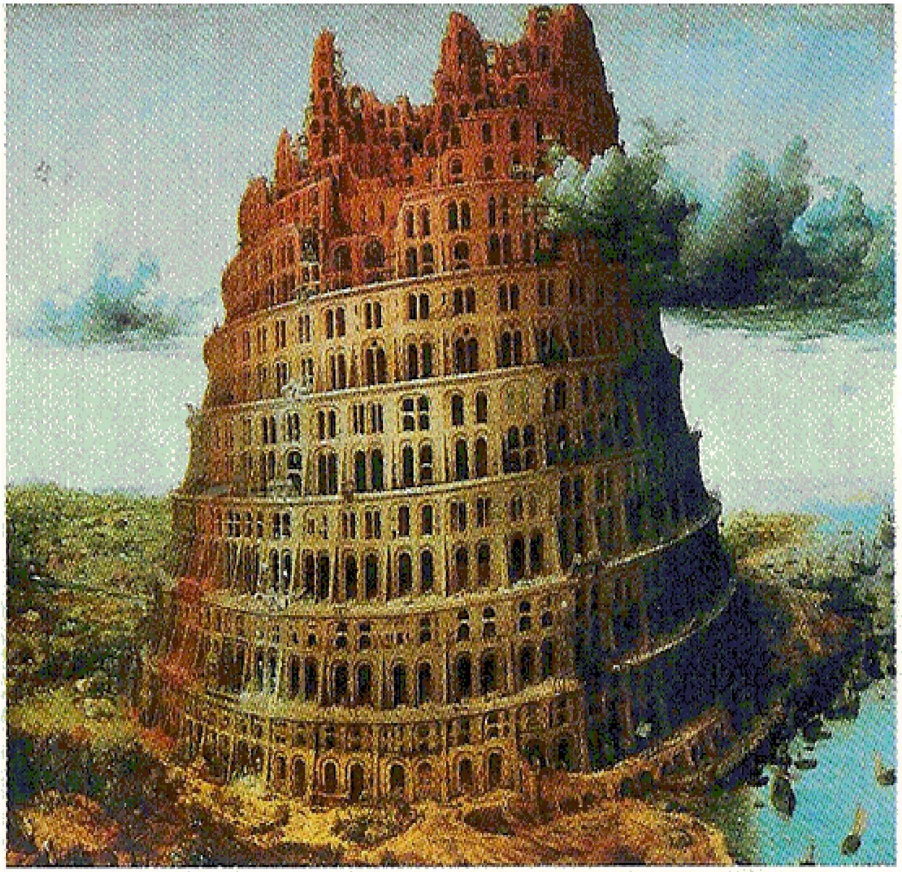
From Wikipedia, the free encyclopedia

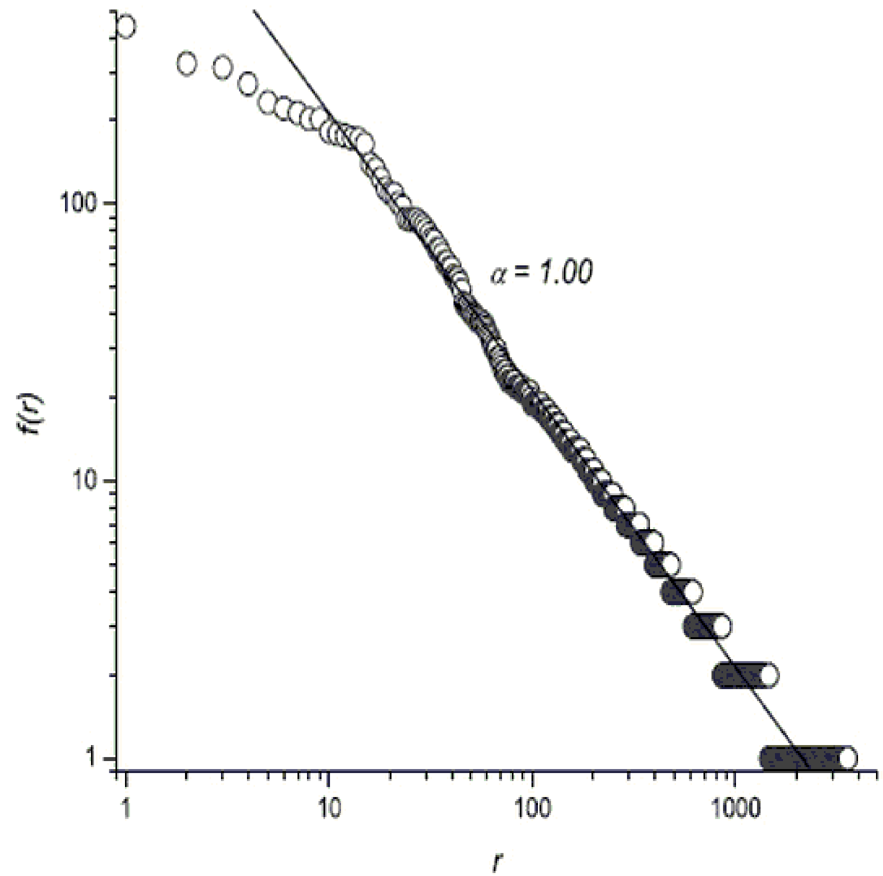
nt
pedia
dia
portal
ges
pedia

File File history File usage



A zebra's stripes, a seashell's spirals, a butterfly's wings: these are all examples of patterns in nature. The formation of patterns is a puzzle for mathematicians and biologists alike. How does the delicate design of a butterfly's wings come from a single fertilized egg? How does pattern emerge out of no pattern?
http://www.sciencedaily.com/releases/2008/06/080619111748.htm

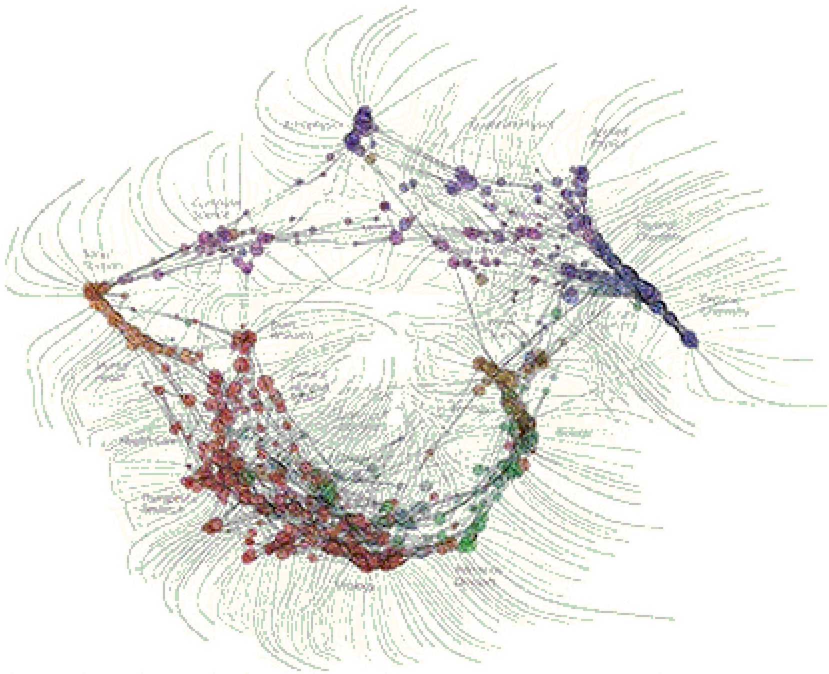




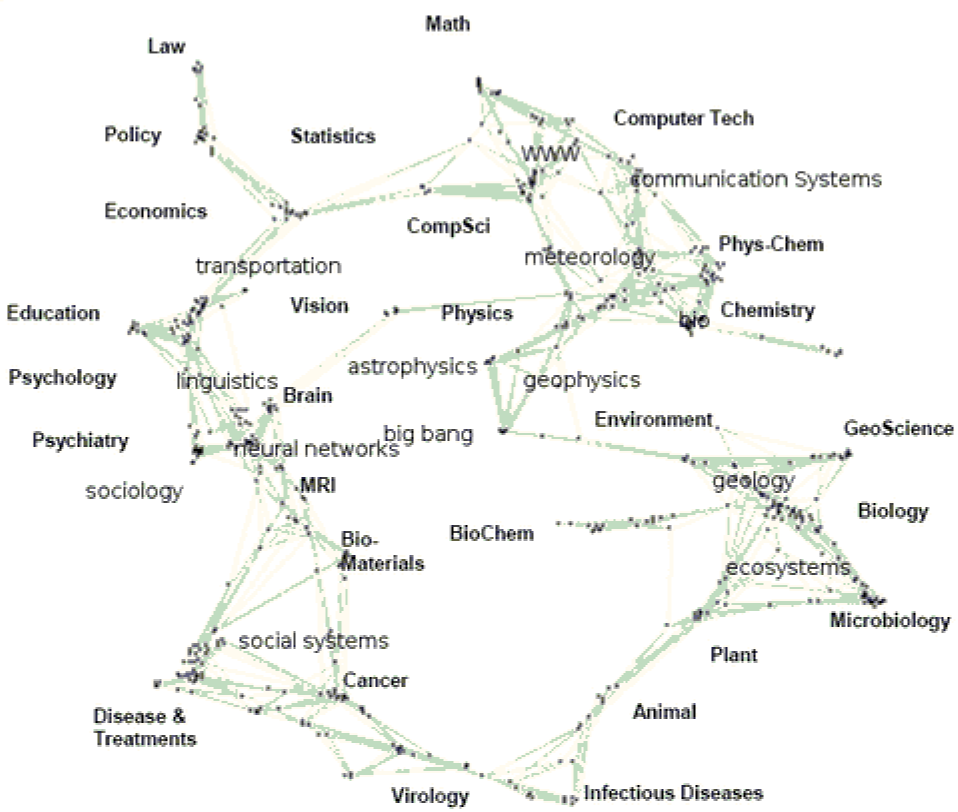
word frequency distributions (Zipf's law)

PZM (Pareto-Zipf-Mandelbrot, parabolic fractal) distributions are observed for the word frequencies of all texts of all languages, all times, any age of author, even the bubbling of babies show the same Pareto-Zipf-Mandelbrot distribution with a slope of 1.00 ...

- scientometrics,
  research and publications :



Research & Node Layout: Kevin Boyack and Dick Klavans (mapofscience.com); Data: Thompson ISI; Graphics & Typography: W. Bradford Paley (didi.com/brad); Commissioned Katy Börner (scimaps.org)



topic map of science with links (click on the subject) to PZM Pareto-Zipf-Mandelbrot (parabolic fractal) distributions.

Self-organizing

DISE

- A system of elements spontaneously forming of well organized structures[de Boer, 1998]
  - Elements are distributed i.e., no single element coordinates the activity
  - Patterns, or behaviors, from random initial conditions.
  - Self limiting, limits its own growth by its actions
  - Universal mechanism for social animals and simple mathematical structures, expected in human society. e.g. the wireless communications industry.
  - Tell-tale signs of self-organization are
    - statistical properties shared with self-organizing physical systems (i.e. Zipf's law, power-law, Pareto principle).
- Emerge from bottom-up interactions, and appear to be limitless in size. Top-down hierarchical networks, which are not self-organizing.
- In economics,
  - Market economy is sometimes said to be [Krugman,1996].
  - Friedrich Hayek coined the term catallaxy as to exchange, to admit in the community and to change from enemy into friend, which is an alternative expression for the word economy, now a new dimension in software design and network architecture [Eymann, Padovan & Schoder, 2000], to describe a "self-organizing system of voluntary co-operation."
  - Central planning is not and less efficient.

Growth Theories Using Centrality

•Degree-based centrality,
•In-degree, out-degree and total degree,
•Google's PageRank algorithm among others such as
  • hub and authority centralities belongs to this group.
•A betweeness centrality describes whether and how frequently a node is part of the shortest paths between pairs of nodes in the network.
•A closeness centrality is defined in terms of the lengths of the shortest paths from a node to the rest of the nodes in the networks.
•***Structure Holes***[Burt, 2005]
  •Structural holes refer to the absence of ties between two parts of a network.
  •Finding and exploiting a structural hole can give an entrepreneur a competitive advantage. Ronald Burt, 1995, 2005], and is sometimes referred to as an alternate conception of social capital
  •Actors with a lot of structural holes (i.e. nonredundant ties) in their network are supposed to hold informational and control advantages that allow them to capitalize from their social networks in ways that others cannot. These people occupy a brokering position. The standard argument is that a network with many structural holes leads to better financial outcomes, greater returns to investment, etc.
  •But it's possible that the standard theory of structural holes is based on an individualistic, Western view of human behavior. That is, it assumes that people adhere to the individualistic principles of Western culture. What happens to people with networks rich in structural holes that live/work in environments that adhere to other principles, such as those of a collectivistic culture?
  •http://orgtheory.wordpress.com/2007/06/19/structural-holes-in-context/

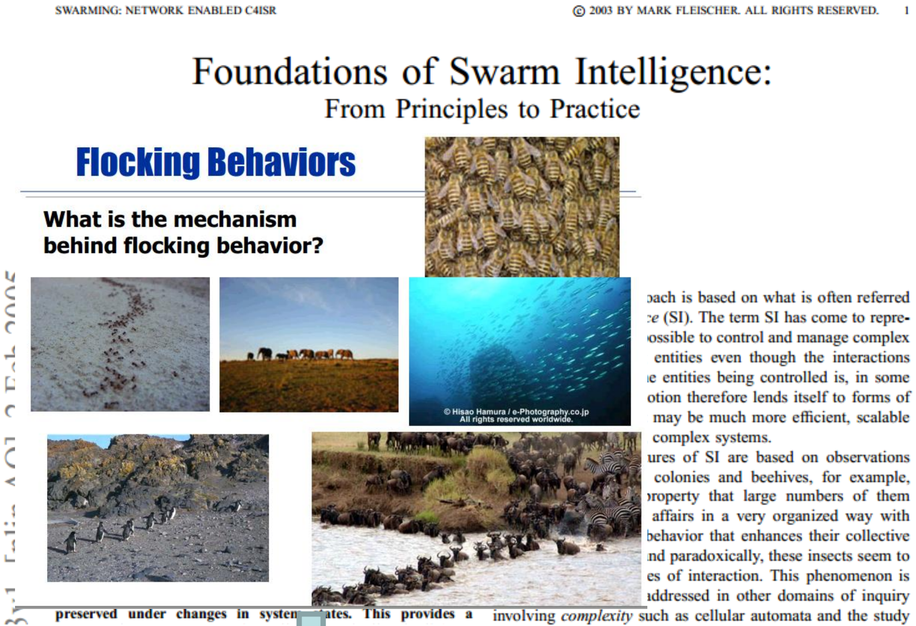
Preferential Attachment (PA) [Barabási & Albert, 1999]

- **The most popular explanation**
  - a new node is connected to a pre-existing one with a probability proportional to the number of links (degree) of the target node
  - any of a class of processes in which some quantity, e.g. wealth or credit, is distributed among a number of individuals or objects according to how much they already have, so that those who are already wealthy receive more than those who are not.
  - 'rich get richer' ,
  - "Yule process",
  - "cumulative advantage",
  - the "Matthew effect".
  - the first application of the process was to grow a random network to a scale-free network[Price, 1976]. Price also promoted preferential attachment as a possible explanation for power laws in many other phenomena
  - Lotka's law of scientific productivity
  - Bradford's law of journal use,
  - Gibrat's law of business or firm growth
  - Zipf's law of city sizes.
- **Successful in predicting the graph structure of the web among others**
- **Problems with PA**
  - As time evolves, new nodes join the network by adding links with a probability proportional to the degree of existing nodes.
  - Higher degree of a node reflects higher relevance or popularity.
  - Earlier nodes tend to have significantly higher degrees than later ones, making it hard for a node which enters late to compete with the already established hubs of the network[Borgs, Chayes, Daskalakis & Roch, 2007].

Pareto Optimal

- Pareto efficient
  - Given an initial allocation of goods among a set of individuals, a change to a different allocation that makes at least one individual better off without making any other individual worse off is called a *Pareto improvement*.
  - An allocation is defined as "Pareto efficient" or "Pareto optimal" when no further Pareto improvements can be made.
- A system that is *not* Pareto efficient
  - implies that a certain change in allocation of goods (for example) may result in some individuals being made better off with no individual being made worse off, and therefore can be made more Pareto efficient through a Pareto improvement.
  - Here *better off* is often interpreted as put in a preferred position, for example, more central or higher degree
- Implications
  - Game theory: <the problem of a coordination failure>
    - The existence of externalities lead to coordination failure and results in Pareto-inferior outcomes.
  - Computer science: <the price of anarchy>
    - Selfish behavior may not achieve full efficiency at the collective level.

http://arxiv.org/pdf/nlin/0502003.pdf


SWARMING: NETWORK ENABLED C4ISR

© 2003 BY MARK FLEISCHER. ALL RIGHTS RESERVED. 1

Foundations of Swarm Intelligence:
From Principles to Practice

...oach is based on what is often referred ...e (SI). The term SI has come to repre- ...ossible to control and manage complex entities even though the interactions ...e entities being controlled is, in some ...otion therefore lends itself to forms of may be much more efficient, scalable complex systems.

...ures of SI are based on observations colonies and beehives, for example, ...roperty that large numbers of them affairs in a very organized way with behavior that enhances their collective ...nd paradoxically, these insects seem to ...es of interaction. This phenomenon is ...addressed in other domains of inquiry involving *complexity* such as cellular automata and the study

preserved under changes in system ...ates. This provides a

- Self-organized to collective better;
- Local, simple communications but achieves Pareto optimal

(http://www.funpecrp.com.br/gmr/year2005/vol3-4/wob09_full_text.htm)

- Use for design armed forces, wireless communications, cellular automata, peer-to-peer networks where one wants to have strong collective intelligence for the whole network/system

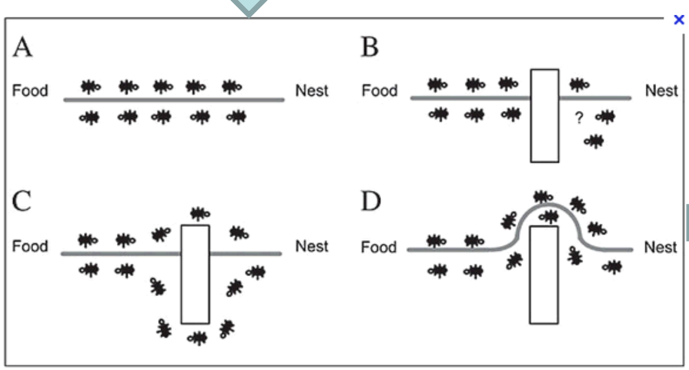


Figure 2. A. Ants in a pheromone trail between nest and food; **B.** an obstacle interrupts the trail; **C.** ants find two paths to go around the obstacle; **D.** a new pheromone trail is formed along the shorter path.

shorter paths have a stronger increment in pheromone

Collaborative Learning Agents

At any given time, we are able to rank the knowledge themes based on its predicted future importance, and distribute themes among stakeholders and social actors.

- Measure the fitness of the whole system. On a theoretic level, we will
  - Hidden Markov Models (HMM) for global optimization with a local learning:

<u>Observations O(t)</u>: Characteristics about a single agent/actor/ that is observable, e.g. measures of single stakeholder's awareness of information using lexical links;

<u>Hidden state *j*</u>, *j=1,…J,* Hidden information that is interesting but difficult to observe directly from data, e.g. stakeholders and regulators can possess different types of competitiveness, reward.

We will also model the predictive relation between lexical links *O(t)* and *hidden* states as a probability density function *bj(O(t)) = b(a(t)=aj|O(t))*. The overall fitness *R(t,aj)* means the total fitness of a complex system up to time *t*. The overall fitness function can be computed recursively.

bj(O(t))

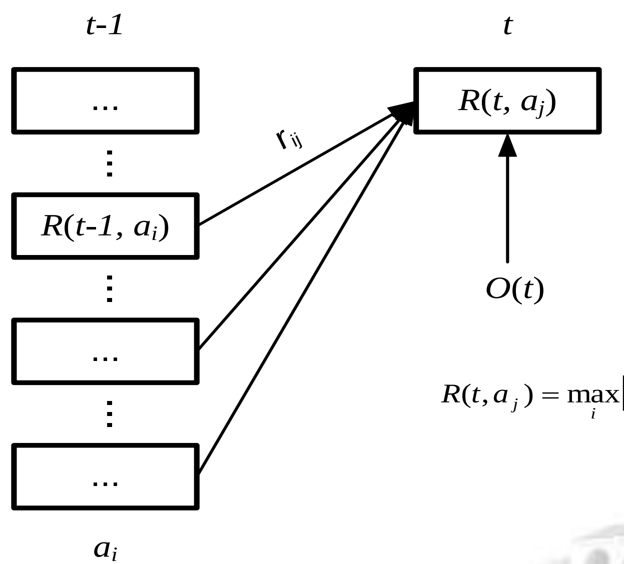
-Measure of reward of a single agent action with the local knowledge of

-e.g. **self-awareness of an individual actor on how different, diversified, anomalous the agent is from others.**

-*R(t,aj)* a global fitness

-Multi-agent systems

Recursion to Compute the Overall Fitness of a System *R(t, aj)*

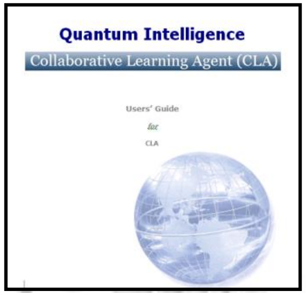


$$R(t,a_j) = \max_i \left[R(t-1,a_i) + r_{ij} \right] + b_j \left[O(t) \right]$$

BACKUP

Table of Contents

Table of Contents



Lexical Link Analysis

- Lexical Link Analysis (LLA) is a form of text analysis
  - A text is represented as a network of lexical terms (e.g. word pairs, bigram) if they are in a community of a word network.
  - Word pairs are further grouped into concepts and themes using large-scale social network community detection algorithms
  - Consequently the importance, impact and evolution of these concepts and themes can be revealed, as well as the crucial relationships among pre-defined categories or automated discovered clusters.
- In a nutshell, LLA is a statistical co-occurrence, bi-gram TAN method for text analysis.
  - *Singlish* (Singapore English mixed English and Chinese)
  - Biological systems within their own symbols for representations.
  - We want to emphasize the connection of LLA's connection to the theories and practices of complex systems and systems of systems, where anticipated benefits of such analysis and presentation are manifested into the concept of System Self-awareness.
- Core focus: Use LLA to automatically discover the concepts and themes in large-scale texts and represent them as dynamic evolving networks over time
  - As a new way to predict the emergence of new information.
  - Discuss the relationship of LLA to complex system theories and network centrality measures.
  - Use cases examine the content of diversified unstructured data, identify new information that might have large impacts and growth potentials in the future.

How LLA Computed

- Read each set of documents.
- Select feature-like word pairs.
- Apply a social network community finding algorithm (e.g. Newman grouping method; Girvan et al. 2001) to group the word pairs into themes. A theme includes a collection of lexical word pairs connected each other.
- Compute a "weight" for a theme for the information of a time period, that is, how many word pairs belong to a theme for that time period and for all the time periods.
- Sort theme weights by time, and study the distributions of the themes by time.
- General questions that LLA usually answers are as follows:
  - Discover themes and topics in the unstructured documents and sort the importance of the themes
  - Discover social and semantic networks of organizations who were involved, compare the two networks to obtain insights to answer the following questions:
  - What were the organizations involved in the *important* themes
  - How do semantic networks suggest more potential collaboration when compared to social networks?

Text Analysis/Mining Tasks

- Named Entity Extraction (NEE)
  - People, place, date, money, etc.
- Text Summary
- Text Categorization
- Text Clustering
- Concept Extraction
- Topic/Theme Extraction
- Text Dynamics: Emergence of New Concepts/Themes Over Time
- Sorting documents, keywords and themes
  - Search

Cognitive Expertise

Informal Name

None

Citation

Carley, Kathleen M. 2002. "Summary of Key Network Measures for Characterizing Organizational Architectures." Unpublished Document: CMU 2002

Description

Measures the degree to which each pair of agents has complementary knowledge, expressed as a percentage of the knowledge of the first agent.

Example : If one person in an organization knows how do perform X but can't do Y. Whereas another individual can do Y but not X, such individuals would rank highly in this measure.

Input : Agent (source) by Knowledge (knowledge) matrix with DataType=binary.

Output : $\Re \in [0,1]$

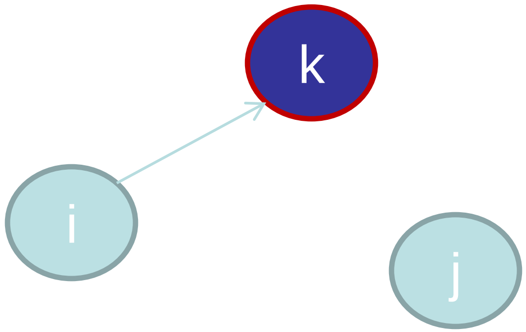
Node Level with Type=agent and DataType=real.

Dyad Level with Type=agent, Target=agent, and DataType=real.

For each pair of agents (i,j) compute the number of knowledge bits that j knows that i does not know. Then normalize this sum by the total number of knowledge bits that agent i does not know.

$$CE_{i,j} = \frac{\sum_{k=1}^{|K|} (\sim AK_{i,k} * AK_{j,k})}{(|K| - \sum_{k=1}^{|K|} AK_{i,k})}$$

$$CE_{i,i} = 0$$



NOTE : The CD output matrix is NOT-symmetric.